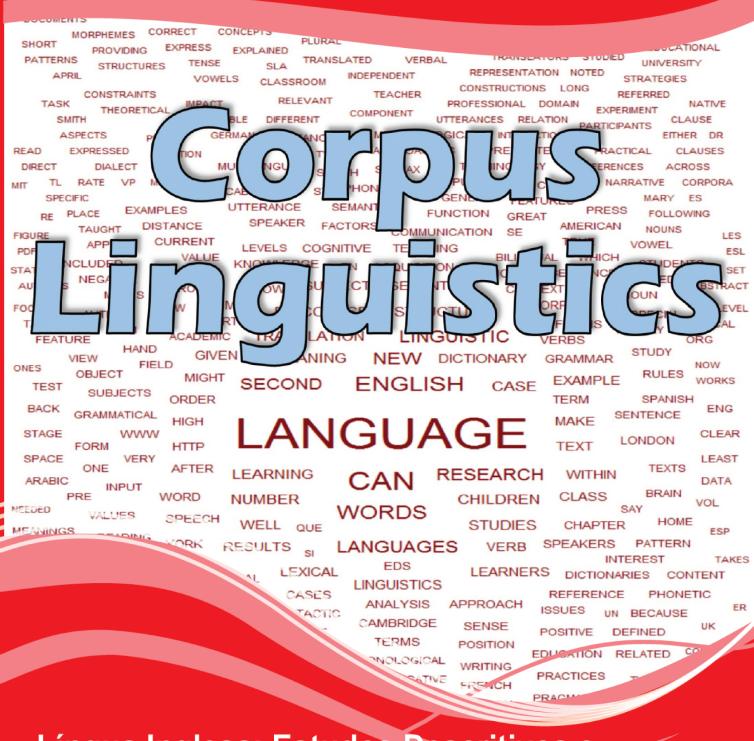
Universidade Federal de Uberlândia Curso de Letras/Inglês



Língua Inglesa: Estudos Descritivos e Linguística de Corpus Integrada à Prática Educativa (PIPE 7) Prof. Dr. Guilherme Fromm

Dados Internacionais de Catalogação na Publicação (CIP) Sistema de Bibliotecas da UFU, MG, Brasil.

F932e	 Fromm, Guilherme, 1968- Estudos descritivos e linguística de corpus integrada à prática educativa (PIPE 7) / Guilherme Fromm 2. ed Uberlândia : Edufu, 2021. 80 p. : il.
	ISBN: 978-65-86084-30-6 (ebook) Guia digital Disponível em: https://doi.org/10.14393/ufu-978-65-86084-30-6 Textos em inglês e Português. Inclui bibliografia.
	 Linguística. 2. Língua inglesa - Ensino à distância. Língua Inglesa - Guias. I. Título.
	CDU 801

PRESIDENTE DA REPÚBLICA Jair Messias Bolsonaro

MINISTRO DA EDUCAÇÃO Milton Ribeiro

UNIVERSIDADE ABERTA DO BRASIL DIRETORIA DE EDUCAÇÃO A DISTÂNCIA/CAPES Carlos Cezar Modernel Lenuzza

UNIVERSIDADE FEDERAL DE UBERLÂNDIA - UFU REITOR Valder Steffen Junior

VICE-REITOR Carlos Henrique Martins

CENTRO DE EDUCAÇÃO A DISTÂNCIA DIRETOR Vinícius Silva Pereira

REPRESENTANTE UAB/UFU Maria Teresa Menezes Freitas

SUPLENTE UAB/UFU Aléxia Pádua Franco

INSTITUTO DE LETRAS E LINGUÍSTICA -ILEEL -UFU DIRETOR Ariel Novodvorski

CURSO DE LETRAS - LICENCIATURA EM INGLÊS E LITERATURAS DE LÍNGUA INGLESA COORDENADOR Rafael Matielo

ILEEL – UFU CONSELHO EDITORIAL

Aléxia Pádua Franco - UFU Bruno Franceschini - UFG Diva Souza Silva - UFU Maria Teresa Menezes Freitas - UFU Simone Tiemi Hashiguti – UFU Stella Esther Ortweiler Tagnin - USP Viviane Cabral Benzegen – UFV

EQUIPE DO CENTRO DE EDUCAÇÃO A DISTÂNCIA DA UFU - CEaD/UFU

ASSESSORA DA DIRETORIA Sarah Mendonça de Araújo

EQUIPE MULTIDISCIPLINAR Alberto Dumont Alves Oliveira Darcius Ferreira Lisboa Oliveira Dirceu Nogueira de Sales Duarte Júnior Gustavo Bruno do Vale Otaviano Ferreira Guimarães

SUMÁRIO

SUMÁRIO	4
FIGURAS	5
TABELAS	7
INFORMAÇÕES	8
SOBRE O AUTOR	9
INTRODUÇÃO	10
AGENDA	11
MÓDULO 1	12
Basic principles of Corpus Linguistics	
ACTIVITY 1	13
ACTIVITY 2	13
Corpus Linguistics?	13
ACTIVITY 3 - SEARCH	
ACTIVITY 4 - DISCUSSION	
Corpus Linguistics in History	
Corpora Typology and Planning	
Key-Concepts	20
MÓDULO 2	29
Corpus Linguistics: Teaching and Learning	30
ACTIVITY 5	30
At school, at home, self-learning	30
ACTIVITY 6 - SEARCH	
ACTIVITY 7 - DISCUSSION	
ACTIVITY 8 - PIPE	38
MÓDULO 3	39
Lexical Analysis Tools	
ACTIVITY 9	40
ACTIVITY 9	
ACTIVITY 9 Compiling a corpus ACTIVITY 10 - CORPUS COMPILATION	42
Compiling a corpus	42 47
Compiling a corpus ACTIVITY 10 - CORPUS COMPILATION AntConc	42 47 47
Compiling a corpus ACTIVITY 10 - CORPUS COMPILATION. AntConc. MÓDULO 4.	42 47 47 67
Compiling a corpus ACTIVITY 10 - CORPUS COMPILATION AntConc	42 47 47 67 68
Compiling a corpus ACTIVITY 10 - CORPUS COMPILATION. AntConc. MÓDULO 4.	42 47 47 67 68 68
Compiling a corpus ACTIVITY 10 - CORPUS COMPILATION AntConc	42 47 47 67 68 68 68
Compiling a corpus ACTIVITY 10 - CORPUS COMPILATION AntConc	 42 47 47 67 68 68 68 68 68 68 68
Compiling a corpus ACTIVITY 10 - CORPUS COMPILATION. AntConc. MÓDULO 4. Project Development. ACTIVITY 11 ACTIVITY 12.	 42 47 47 67 68 68 68 68 68 69
Compiling a corpus ACTIVITY 10 - CORPUS COMPILATION AntConc	42 47 47 67 68 68 68 68 68 69 71

FIGURAS

	Availability	Page
Figure 1	Source: YAMAMOTO, Márcio Issamu. VoBLing: vocabulário bilíngue de	
	linguística, português-inglês, direcionado por corpus. 2020. 214 f. Tese	
	(Doutorado em Estudos Linguísticos) - Universidade Federal de Uberlândia,	14
	Uberlândia, 2020. Disponível em: http://doi.org/10.14393/ufu.te.2020.682.	
Figure 2	Source: from the author.	15
Figure 3	Source: from the author.	16
Figure 4	Source: from the author.	16
Figure 5	Source: from the author.	17
Figure 6	Source: from the author. Wordlist of the Linguistics Corpus.	21
Figure 7	Statistics. Source: from the author.	22
Figure 8	Keywords for the Sociolinguistics subcorpus. Source: from the author.	23
Figure 9	Concordance lines for the word VALENCE in the Linguistics Corpus. Source:	24
	from the author.	24
Figure 10	Text tab for the word VALENCE; line 16 (valence alternation) chosen. Source:	25
	from the author.	-
Figure 11	Collocates for the word VALENCE. Source: from the author.	26
Figure 12	Concordance lines with the words VALENCE and SUFFIXES. Source: from the	27
	author.	
Figure 13	Cluster tab for the word VALENCE. Source: from the author.	28
Figure 14	Concordance lines for the cluster verb root valence. Source: from the author.	28
Figure 15	Registration corner. Source: COCA. Available at: <u>https://www.english-</u>	30
	<u>corpora.org/coca/</u>	
Figure 16	COCA's upper menu. Available at: <u>https://www.english-corpora.org/coca/</u>	31
Figure 17	COCA's search display. Source: COCA. Available at: <u>https://www.english-</u>	32
	<u>corpora.org/coca/</u>	52
Figure 18	COCA's query result. Source: COCA. Available at: <u>https://www.english-</u>	32
	corpora.org/coca/	
Figure 19	COCA's concordance lines for the query at the Internet. Source: COCA.	33
	Available at: https://www.english-corpora.org/coca/	55
Figure 20	Expanded concordance line for the first example. Source: COCA. Available at:	33
Figure 21	https://www.english-corpora.org/coca/	
Figure 21	Word and Phrase main page. COCA platform. Source: COCA. Available at:	35
Figure 22	https://www.english-corpora.org/coca/ Frequency query for Linguistics. Source: COCA. Available:	
Figure 22		36
Figure 23	https://www.english-corpora.org/coca/ Word LINGUISTICS description. Available at: https://www.english-	
Figure 25	Word LINGUISTICS description. Available at: <u>https://www.english-</u> corpora.org/coca/	37
Figure 24		
Figure 24	Menu to get to new screens. Source: COCA. Available: <u>https://www.english-</u>	37
Figure 25	<u>corpora.org/coca/</u> Text insertion. Available at: <u>https://www.english-corpora.org/coca/</u>	27
Figure 26		37
Figure 20 Figure 27	Text analysis. Available at: <u>https://www.english-corpora.org/coca/</u> Corpora do Português, COCA plataform. Source: COCA. Available:	38
i iyure 27	Corpora do Português, COCA plataform. Source: COCA. Available: http://www.corpusdoportugues.org/x.asp	38
Figure 28	WordSmith Tools' main menu. Source: from the author.	40
Figure 28	WordSmith Tools main menu. Source: from the author.	
Figure 29 Figure 30		41
Figure 31	AntConc's main menu. Source: from the author.	42
Figure 32	Google: query for the word <i>Prosody</i> . Source: from the author. File Prosody: <i>Rhythms and Melodies of Speech</i> first page. Source:	43
i igule 52		44
Figure 33	https://arxiv.org/ftp/arxiv/papers/1704/1704.02565.pdf.	A A
Figure 33	Path to open the Notepad program. Source: from the author.	<u>44</u> 45
Figure 34	File selected. Source: from the author.	
i iyule JJ	File copied to the Notepad program.	45
	Header insertion. Source: from the author.	46
Figure 36	Coving the file of File 1. Coving theme the surther	
Figure 36 Figure 37	Saving the file as File 1. Source: from the author	46
Figure 36 Figure 37 Figure 38	AntConc. Opening a directory in the WordList tool. Source: from the author.	47
Figure 36 Figure 37		

Figure 41	AntConc. Tool Preferences menu. Source: from the author.	50
Figure 42	Word List with Stoplist, Prosody. Source: from the author.	51
Figure 43	Loading the Lemma List. Source: from the author.	52
Figure 44	Word List with Stoplist and Lemma list. Source: from the author.	53
Figure 45	Saving file in AntConc. Source: from the author.	54
Figure 46	File saved from AntConc's Word List. Source: from the author.	54
Figure 47	Concordance lines for the word PRAGMATIC. Source: from the author.	55
Figure 48	File View for the first entry of the word PRAGMATIC. Source: from the author.	55
Figure 49	Concordance Plot for the word PRAGMATIC. Source: from the author.	56
Figure 50	Concordance lines for the word PRAGMATICS. Source: from the author.	57
Figure 51	Kwic sort configurations for the word PRAGMATICS. Source: from the author.	57
Figure 52	Searching for the specific word SPEAKER in the WordList. Source: from the author.	58
Figure 53	Collocates for the word SPEAKER. Source: from the author.	59
Figure 54	Collocates for SPEAKER + HEARER. Source: from the author.	60
Figure 55	Clusters for the word SPEAKER. Source: from the author.	61
Figure 56	Clusters for the word SPEAKER. Source: from the author.	62
Figure 57	SPEAKER'S UTTERANCE clusters. Source: from the author.	63
Figure 58	AntConc's Tool Preferences and Keyword List. Source: from the author.	64
Figure 59	Keyword List from the PRAGMATICS corpus. Source: from the author.	65
Figure 60	Concordance lines for the word DISCOURSE. Source: from the author.	66
Figure 61	Concordances lines for the combination PRAGMATICS + IS. Source: from the author.	70
Figure 62	Concordance for the combination PRAGMATICS + :. Source: from the author.	70
Figure 63	Main consulting page of VoTec (in English), Linguistics area. Source: from the author.	71
Figure 64	VoTec researcher's first page. Source: from the author.	72
Figure 65	Creating the term PROSODY. Source: from the author.	72
Figure 66	Inserting a context for PROSODY. Source: from the author.	73
Figure 67	Inserted contexts for PROSODY. Source: from the author.	74
Figure 68	Second page for the term PROSODY. Source: from the author.	74
Figure 69	Data tab. Source: from the author.	75
Figure 70	Distinctive Traces tab. Source: from the author.	75
Figure 71	Semantics tab. Source: from the author.	76
Figure 72	Equivalent Term tab. Source: from the author.	76
Figure 73	Cross-reference terms tab. Source: from the author.	77
Figure 74	Encyclopedic Information tab. Source: from the author.	77
Figure 75	Final Concept and Definition tabs. Source: from the author.	77
Figure 76	Term PROSODY in VoTec's page (just English). Source: from the author.	78
Figure 77	Term GRAMMAR in the VoTec. Source: from the author.	78

TABELAS

Table 1. Collaborative Corpus of Linguistics typology	
Table 2. Features related to COCA.	
Table 3. Query possibilities using the COCA interface.	35
Table 4. Corpus compilation	

INFORMAÇÕES

Prezado(a) aluno(a),

Ao longo deste guia impresso você encontrará alguns "ícones" que lhe ajudará a identificar as atividades.



Fique atento ao significado de cada um deles, isso facilitará a sua leitura e seus estudos.

Destacamos alguns termos no texto do Guia cujos sentidos serão importantes para sua compreensão. Para permitir sua iniciativa e pesquisa não criamos um glossário, mas se houver dificuldade interaja no *Fórum de Dúvidas*.

SOBRE O AUTOR

Guilherme Fromm é professor de Língua Inglesa do Instituto de Letras e Linguística da Universidade Federal de Uberlândia. Cursou História e Letras (Alemão/Português) na graduação. Especializou-se em Tradução (Inglês/Português), cursou o mestrado na área de Linguística, o doutorado na área de Letras - Estudos Linguísticos e Literários em Inglês e o pós-doutorado na área do Léxico. Toda sua formação acadêmica se deu na Universidade de São Paulo, com exceção do pós-doutorado, realizado na UFSCar.

Atuou como professor de inglês em cursos livres por quinze anos e há dezenove é professor universitário, nas áreas de Linguística e Língua Inglesa. Diretor das revistas Domínios de Lingu@gem (http://www.seer.ufu.br/index.php/dominiosdelinguagem), GTLex (http:// www.seer.ufu.br/index.php/GTLex/index) e da série de livros e-Classe (pela EDUFU) desde o seu início. É autor de artigos e capítulos de livros e organizador de livros. Tem experiência na área de Linguística, atuando nos seguintes temas: Ciências do Léxico (Lexicologia, Lexicografia, Terminologia, Terminografia), Linguística de Corpus, Ensino de Língua Inglesa e Tradução.

INTRODUÇÃO

Olá! Seja bem-vindo.

Com este guia, trabalharemos uma disciplina diferenciada dentro do curso PARFOR/ Inglês. O objetivo geral é aprender a descrever língua, ao invés de só ensiná-la. Para tanto, adotaremos a abordagem e a metodologia da Linguística de *Corpus*.

Embora a ênfase do nosso trabalho, aqui, seja a descrição, também mostraremos como ferramentas da Linguística de *Corpus* podem ser trabalhadas na sala de aula para o enriquecimento da experiência de aprendizagem do aluno.

Como trabalho prático para o PIPE, compilaremos um *corpus* e prepararemos um plano de aula detalhado usando as várias possibilidades de atividades aqui expostas.

AGENDA

WEEK	MODULES	STUDY DEVELOPMENT	EVALUATION
Weeks 1 and 2	Module 1 Basic principles of Corpus Linguistics	Activity 1 – chat with the tutor. Activity 2 – video class. Activity 3 – Corpus Linguistics search. Activity 4 – Forum discussion	Activity 4 Score: 10 points.
Weeks 3 and 4	Module 2 Corpus Linguistics: Teaching and Learning	Activity 5 – video class. Activity 6 – search using COCA. Activity 7 – Forum discussion Activity 8 – PIPE. Search and describe a site that uses corpora and tools to analyze language.	Activity 7 Score: 5 points. Activity 8 Score: 10 points.
Weeks 5 and 6	Module 3 Lexical Analysis Tools	Activity 9 – video class. Activity 10 – Corpus Compilation. You're going to compile a study corpus.	Activity 10 Score: 10 points.
Weeks 7 and 8	Module 4 Project Development	Activity 11 – video class. Activity 12 – PIPE. You're going to prepare a class plan project, using the Corpus Linguistics approach.	Activity 12 Score: 15 points.

MÓDULO 1

Basic principles of Corpus Linguistics

Basic Contents

- What is Corpus Linguistics?
- What to consider?
- Corpus Linguistics in History.
- Corpora typology and planning.
- •Key-concepts: frequency, keywords, concordances, collocations, clusters/n-grams.

Objectives

- Delimitate the activity field.
- •Know that the practice is old, but the way we work today is new.
- Decide the ways you implement a research.
- •Know the importance in corpus design and how to identify it.
- Understand the basic concepts related to Corpus Linguistics.

MÓDULO 1

BASIC PRINCIPLES OF CORPUS LINGUISTICS

ACTIVITY 1

Chat with your tutor. Take a look at the site (AVA) the date and schedule of this chat.

ACTIVITY 2

Video class, module 1. Watch the professor's hints about the subjects that are going to be worked in this module.

Corpus Linguistics?

First, we need to delimitate the field we are studying. What is Corpus Linguistics (CL from now on)? Let us see a definition provided by the CASS Centre:

Corpus linguistics, broadly, is a collection of methods for studying language. It begins with collecting a large set of language data – a corpus -, which is made usable by computers. Corpora (the plural of corpus) are usually so large that it would be impossible to analyse them by hand, so software packages (often called concordancers) are used in order to study them. It is also important that a corpus is built using data well matched to a research question it is built to investigate. To investigate language use in an academic context, for example, it would be appropriate for one to collect data from academic contexts such asacademic journals or lectures. Collecting data from the sports pages of a tabloid newspaper would make much less sense.



An important idea we must consider is that CL, nowadays, means the usage of a computer. This is a kind of **empiric** (based on facts) work.



ACTIVITY 3 - SEARCH

What about you finding other concepts of CL available on the Internet? Do a search using Google (or other searcher) about the term. What have you found interesting about the theme?



ACTIVITY 4 - DISCUSSION

Discuss with the other students, using the AVA forum, which concepts about CL you considered interesting. Try to identify possible approaches for this subject. The participation in the discussion will be evaluated.

What to consider?

When working with research, you must have in mind that there's a division among theory

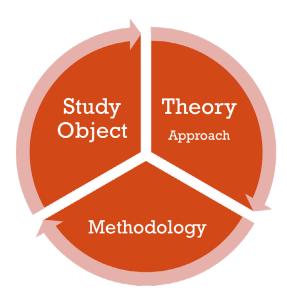


Figure 1. The process of analyzing science.

Just in our area, we have a lot of study fields:

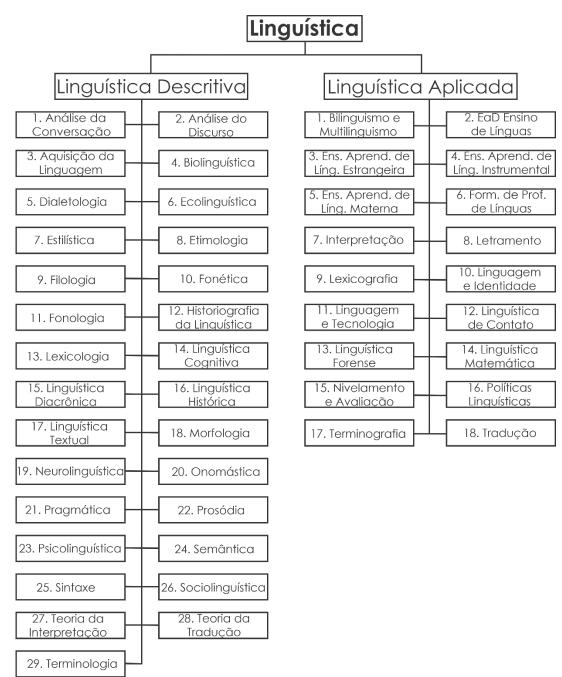


Figure 2. Linguistics Domain Tree.

Basic principles of Corpus Linguistics

Theories:

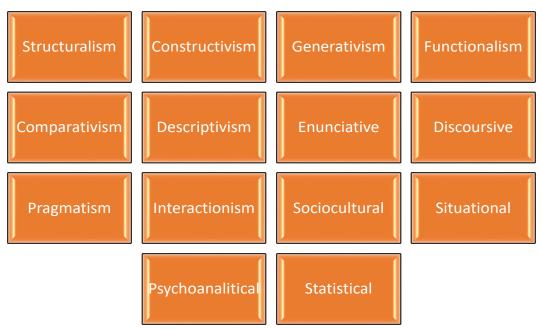
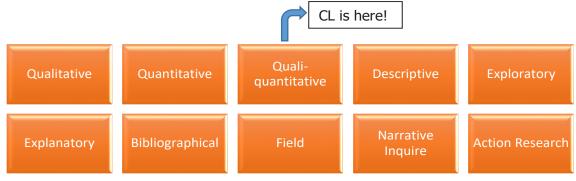


Figure 3. Some linguistic Theories.

And methodologies (this is a tiny example, there are more methodologies):





Corpus Linguistics in History

There's nothing new about collecting corpora for studies. Researchers have been doing this for ages. According to Berber Sardinha:

Havia corpora antes do computador, já que o sentido original da palavra corpus é corpo, conjunto de documentos (conforme o dicionário Aurélio). Na Grécia Antiga, Alexandre, o Grande definiu o Corpus Helenístico. Na Antiguidade e na Idade Média, produziam-se corpora de citações da Bíblia.

Many corpora in English language have been developed since the advent of the computer age. From the early modern corpora (like Brown Corpus), through those who were a mark (like BNC) up to the most contemporary ones (like COCA), the size differ a lot. The Brown Corpus had 1 million words, the BNC had 100 million, and the COCA has 450

PAR

http://

PARE

million (up to the moment). There are corpora that oversize one billion words, like the GloWbe (Global Web-Based English), with 1.9 billion. Take a look here (http://corpus. byu.edu/) and compare sizes.

What is new about the way we're studying today is the approach we have: we develop a research project, from the scratches to the results, using a computer. That's the reason we can say that **Corpus Linguistics is as an approach as well as a methodology** to be used in linguistic studies.

We can't mention the history of Corpus Linguistics without Chomsky. When developed the Generative Grammar first concepts (with the ideas of competence and performance), he criticized a lot the researchers who were using corpora for analysis.

The Generative Grammar and other Chomskyian concepts are particularly important in Linguistics history. Let's learn about these concepts? Search the net and write some of them down.

Not all kinds of theories and methodologies are compatible. When we use the Generativist theory, for example, we can't use the CL methodology, since they're not compatible. On the other hand, although CL seeks the idea how you perform language (we study the way languages are really used), the computational data set of CL lexical analysis programs are based on the Generative Grammar (the theoretical competence of a determined language). A paradox!



Generativism competence



Be careful with the choices!



Corpus Linguistics performance

Computational Paradox

Figure 5. Comparing Generativism and Corpus Linguistics.

Corpora Typology and Planning

Corpus compilation is something that must be carried out very carefully. Imagine you've compiled a study corpus for a project and you discover, later, that the data you have collected is not big enough or not balanced enough throughout the areas you wish to study, or the sources you used are not trustworthy. The whole set of results that come from the corpus, in this case, can be invalidated.

See the table below. It's a summary of the way a corpus was organized.

Língua	Bilíngue (Inglês e Português)
Modo	Escrito (textos acadêmicos: artigos científicos, dissertações e teses)
Data de Publicação	Sincrônico (levantamento realizado entre 2010 e 2014)
Seleção	Amostragem, Estático
Conteúdo	Especializado (Linguística)
Autoria	Falantes nativos/não nativos (inglês e português), individual/coletivo
Disposição Interna	Comparável
Uso na Pesquisa	Estudo (Análise terminológica/terminográfica)
Tamanho	Grande (mais de 10 milhões de palavras)
Nível de Codificação	Sem cabeçalhos, sem etiquetas

Table 1. Collaborative Corpus of Linguistics typology. Source: Fromm, Yamamoto (2013)

Let's take a look at each of these possibilities in corpus designing?

Language

Depending on your project, you can analyze one, two or multiple languages (for a contrastive work, for example). The example from table 1 works with two languages (English and Portuguese).

Sources

Basically, we have two sources corpora are compiled from: written sources or oral sources. Written corpora are very easily found on the Internet, can be scanned from books, etc. Oral corpora are more difficult to be worked with: the programs used for Corpus Linguistics don't analyze the sounds of recordings; we must transcribe the sounds into a written file first, and then analyze them.

Nevertheless, other combinations can be arranged. Beilke¹, for examples, intends to use the texts from gravestones to analyze the Pommern dialect used in Brazil by some

¹ BEILKE, Neubiana Silva Veloso. Pommersche Korpora: uma proposta metodológica para compilação de corpora dialetais. 2016. 285 f. Dissertação (Mestrado em Estudos Linguísticos) - Universidade Federal de Uberlândia, Uberlândia, 2016. Disponível em: http://doi.org/10.14393/ufu.di.2016.426

communities that immigrated from Germany and Poland. It's an unusual source of a written corpus.In addition, she also intends to analyze traditional written (from clerical sources) and oral (from interviews) corpora as sources.

Time

You can try to compile a diachronic corpus or a synchronic one. A diachronic corpus is much more difficult to be prepared, because it generally involves scanning work.

Selection

Great corpora of a language, like the COCA, are called general or reference corpora. The corpus you prepare for your research, much smaller, is called sample or study corpus. If you prepare your corpus and finish it, it's called a static corpus. If the corpus is continuously changed, we name it dynamic corpus. All these previous possibilities imply the idea of **balance**: the amount of texts or numbers of words must be very well distributed among the corpora genres or subsections.

Content

A general corpus must contain, theoretically, all the traces of a language; it's very difficult to design one, since a deep previous study must be donebefore you start collecting texts. A study corpus contains texts that are related to the study you're developing in your research.

Authorship

The texts you collect for your corpus may be written/spoken by native or non-native speakers of a language. These texts can also be produced individually or in a group (more than one author/speaker).

Internal distribution

For bilingual or multilingual works, the corpora you prepare are parallel or comparable. For example, if you compile texts in English and their translations into Portuguese, these corpora are *parallel*. On the other hand, you can choose an area for your study and take texts you find about that area in each language; they're not the same texts, but belong to the same area – in this case, they are *comparable*.

Size

Berber Sardinha proposes a table to classify the size of a corpus. It's a little bit tricky, however, since nowadays anyone can download hundreds of texts in minutes using the Internet. Let's say that, obeying the other parameters we comment here, the bigger the

size of the corpus the best it is to represent the language (specialized or not).

Codification level

For more specific kinds of studies, the corpora may contain tags (like the morphossyntatic ones, in which each word of a corpus is classified according to its word classes or syntactic

Lingüística de Corpus, from Berber Sardinha, is a very important book in Corpus Linguistics area in Brazil. It contains all the basic research principles for the area. Read it!

Key-Concepts

There are some basic concepts involving the Corpus Linguistics approach that we must always have in mind. We're going to present some of them using screens of the WordSmith Tools (version 6²) program. For our work with lexical analysis software in module 3, however, we're going to work with AntConc, since it's a free software.

Frequency

Frequency is the basis of the Corpus Linguistics. Remember we're working in an empiricist way, so the facts we analyze are based on numbers and statistics. The first kind of a work you do with your corpus is to process it through a wordlist tool. See picture 1 for an example of a wordlist.

² The lexical analysis suite is already in version 8.

] ile <u>I</u>	dit	View	Compute	<u>S</u> ettings	Windows	<u>H</u> elp					
N	-			Word		Freq.	%	Texts	%	L	ę
1				THE		35.103	5,17	1.557			
2				OF		38.380	3.31	1.558			
4				AND		55.623	2.48	1.557	99,81		
5				A		29.708	2.35	1.555	99,68		
6				IN		01.361	2.24	1.556			
7				TO		42.813	2.02	1.557			
8				IS		12.029	1.16	1.549			
9				ТИЛТ		74.177	1.02	1.100			
10				THAT		6.795	1,00 0,88	1.536			
11				E		36.447 26.330	0,88	1.388			
12				FOR		08.442	0,84	1.544			
13				AS		06.614	0.78	1.544			
14				S		77.989	0.66	1.404			
15			1.4			53.163	0.57	1.477			
16			L/	ARE		49.951	0.56	1.534			
17				THIS		43.674	0.54	1.533			
18				BE		42.669	0.53	1.529			
19				N		38.344	0.52	1.072			
20				Т		37.151	0.51	1.193			
21				WITH		34.872	0.50	1.548			
22				ON		32.258	0.49	1.537	98,53		
23				R		29.867	0,48	1.155	74,04		
24				Ĩ		24.325	0,46	1.525	97,76		
25				OR		15.875	0,43	1.528			
26				BY		14.474	0,43	1.540			
27				0		12.602	0.42	766			
28				NOT		03.454	0.39	1.506			
29				C		93.841	0,35	1.223			
30				AN	9	93.170	0.35	1.522	97,56		
31				FROM		90.602	0,34	1.531	98,14		
32				WHICH		39.464	0,33	1.504	96,41		
33				D		31.837	0,31	1.261	80,83		
34				P		31.438			76,22		
35				Μ		79.686	0,30				
36				HAVE		75.897	0,28				
37				G		72.238	0.27				
38				THEIR		58.306	0.25				
39				AT		6.332	0,25				
40				WAS		6.311	0.25				
41				THEY		6.231	0,25				
42	1			CAN	6	54.349	0.24	1.495	95,83		
requer	10V	alphabet	tical statis	tics filenames	notes				10.04		

Figure 6. Wordlist of the Linguistics Corpus.

This first screen of the Wordlist tools emphasizes the frequency (tab 1). As you can see, the article THE is the first one. It appears 1.385.103 times in the corpus texts; it represents 5.17% of all the words in the texts; it appears in 1.557 texts, which represent 99.81% of the texts of the corpus.

As you can see, the grammar words (prepositions, articles, pronouns, etc.) are much more frequent than content words (nouns, verbs, adverbs, adjectives). Usually, the article THE is the most frequent word in English.

If we want more information about the corpus, we can go to the statistics tab (1). Take a look at figure 7:

			WordLi	st									- 6
<u>E</u> di	t View Compute Settings Windows Help												
			tokens (running words) in text	tokens used for word list	sum of types (distinct	type/token		STTR	STTR mean	word length	s mean (in	5	ş mean (in
N	text file	file size	words) in text	word list	e (disunce	ano (TTR)	standardised TTR		e tword lengt		(m		(m
1	Overall	263.638.512	26.805.858	25.355.888	344.032	1.36	35.79	69.91	1.00 4.7	2 3.11	1.35 21	4 1.22	.86 13.6
2	Corpus de Linguística - Inglês.Ist	9.008.217	1.789.089	1.764.185	6.008	0.34	6,90	87.57	1.00 1.1	1 0,60	4.43 962	. 19.9	1 1.76
3	A Rationale For Foreign Language Education.txt	27.802	2.124	2.024	651		40.05	42,39	1.00 5.3	1 2,90	129 15,	6 15,0	1 2 0
4	ASSESSING SPEECH PRODUCTION IN ENGLISH AS A FOREIGN.txt	685.100	51.576	48.706	3.133		32.89	64.50	1.00 5.3	0 3.14	2.70 17.	9 21.5	1 68.
6	Benefits of Being Bilingual.txt	38.762	2.862	2.680	837	31.23	41.90	41.08				5 12.1	1 2.6
6	CLASS PLAN, PRACTICE AND REPORT.txt	539.524	42.857	40.387	4.176		35,20		1.00 4.8				1 40.
7	Cognitive Benefits of Learning Languages.txt	29.454	2.275	2.265	680		38.10		1.00 5.2			1 10.9	1 22
8	Contrastive Analysis and Native Language Identification.txt	78.888	6.317	5.364	1.177		33.50		1.00 4.8				1 03
9	Critical reading in L2.txt	399.610	31.880	30.155	2.974	9,86	34.67	61.47					1 30.
	EMERGING TECHNOLOGIES.txt	61.044	4.583	4.320	1.302		45.75		1.00 5.2				1 4.3
1	ENGLISH AS A FOREIGN LANGUAGE STUDENTS.txt	605.722	49.097	48.075	4.867	10,12	40.09		1.00 4.8				1 48.
2 F	irst Language Communication in the Second Language Classroom A Valuat	107.408	8.404	7.496	1.516		36,36		1.00 4.9				1 7.4
13 14 C	First Language Use in Second Language Literacy Development.txt	37.374	3.101	3.084	749		33.83	50,56	1.00 4.7				1 6.0
	Suidelines for Starting an Elementary School Foreign Language Program.txt	21.872	1.589	1.523	511		37.40		1.00 5.5				1 1.5
5	How native-like is non-native language processing.txt	84.774	6.296	5.613	1.286		37.12		1.00 5.3				1 5.6
6	Informed Use of the Mother Tonque in the English Language Classroom.txt	79.622	6.336	5.861	1.335		37.97		1.00 4.9				1 5 8
7 8	Learning difficulty, L2 proficiency, and implicit and explicit knowledge.txt	397.442	30.282	27.410	3.003		31.10	63.20					1 🖾
3	Native language interference in learning a second txt	66.038	5.446	5.018	967	19.27	30,94		1.00 4.8				1 5.0
	Oral Development in English as a Foreign Language.txt	826.916	64.832	62.364	4.082		30,16		1.00 4.9				1 62.
0	Strategies in Learning and Using a Second Language.txt	29.038	2.332	2.298	741		39,95		1.00 5.0				1 2.2
2	TEENAGERS' MOTIVATION IN SECOND LANGUAGE ACQUISITION.txt	384.600	28.459	26.967	3.839		36.91		1.00 5.0				1 22
	The Benefits of Second Language Study.txt	76.892	5.455	5.076	1.198		38.70		1.00 5.4				1 5 2
	HE EFFECTS OF AGE AND PROFICIENCY ON VERBAL MORPHOLOGI	388.238	29.736	27.519	2.661	9,67	29,54		1.00 5,1				1 27.
14	The influence of training and instruction on the production of verbs.txt	284.368	22.271	20.009	2.679		34.75		1.00 4.8				1 20.
26	The Internet and Foreign Language Education.txt	49.088	3.786	3.705	1.029		41.60		1.00 5.2			0 10.6	1 3.7
7	THE MOTHER TONGUE IN THE CLASSROOM.txt	40.436	3.182	3.072	870		40,23		1.00 5.0			8 14.6	1 3.0
18	The Relation between High School Study of Foreign Languages.txt	27.758	2.186	2.095	517	24.68	30.55		1.00 5.0				1 2.0
29 29	THE RELATIONSHIP BETWEEN ELEMENTARY SCHOOL FOREIGN.txt	644.424	47.586	44.701	3.755		33.04		1.00 5.4				1 44.
10	THE TEACHING AND LEARNING OF READING AT UERN.txt	385.360	28.679	27.372	3.201	11.69 28.16	34.64	62,39					1 27.
11	Using the mother tongue to promote noticing.txt	37.698	3.068	2.983	840		38.87		1.00 4.9				1 2.9
2	WORKING MEMORY CAPACITY AND ATTENTION TO.txt	643.032	48.523	43.893 3.525	5.003		35.27	61.80					1 43.
33	ROLE OF MOTHER TONGUE IN LEARNING.txt	50.796 28.276	3.945 2.235	3.525	886 693		32.93		1.00 5.0			7.11.0	1 3.5
4	Process of L1 Acquisition.txt The Impact of Mother Tongue on Students' Achievement in.txt	28.276	2.235	2.109	1,112		39.55 31.30	42.74				3 14.8	1 2.1
35		75.520	2.875	2.815	1.112		31,30		1.00 5.0				1 5.4
16	a global perspective.txt acquisition of chinese.txt	38.474	2.875	2.815	1.683	31.26	41.40 36.83	41,44 54,93				7 14.3	1 2.8
7		861.510	63.327	59.929	6.075		30.83	58.03					1 59.4
18	aronin hufesein.txt aspects of multilingualism.txt	119.696	9.449	59.929 9.106	2.506		39.52	58,03 46,80				1 15.4	1 9.1
39	aspects of multilingualism.txt Benefits of linguistic diversity and multilingualism.txt	153.836	9.449	9.106	2.506		47.67	46.80					1 9.1
40	Benefits of inquistic diversity and multilingualism.txt beyond bilingual education.txt	64.676	5.046	4,776	2.092		36.16	52.61				5 14.4 B 20.0	1 11.
41	Bi multilingualism Lissues and Concerns tyt	26 314	5.046	4.770	1.120		37,40	52,61			200 23,		1 4.7
	_ 1	<i>(</i>)))	1 405	1 400	D/0								

Figure 7. Statistics

In the statistics tab (2) we can find: the name of the text that belong to the corpus (text file);the size of the corpus in bytes: approx. 263.63 Mb; the size of the corpus in total words (or tokens): 26.805.858 words; the size of the corpus used for the wordlist: 25.355.88 words³; the amount of distinct words (types⁴): 344.032 different words; the token/type ratio⁵; etc.

When we're talking about Corpus Linguistics, we're talking about computers. This also means we're talking about statistics, programming, Computational Linguistics, mathematics, etc. If you want to get deeper in this area, studying these topics is surely a good idea.

³ By default, the programs don't count numbers (as words) for analysis.

⁴ As you can see in fig. 1, the article THE appears 1.385.103 times (tokens) in the corpus, but it's just one word (type). ⁵ You divide one by another. The idea is the following: the bigger the number of the ratio, the lexically denser the corpus is. Theoretically, a denser text is more difficult to be read.

Keywords

With the keywords tool, we can find out the words that are key to one area of study. In figure 8, we have the keyword list for the Sociolinguistics subcorpus of the Linguistics corpus.

Ν	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P	L
1	LANGUAGE	5.249	1,03	??	22.329	0,02	30.930.87	0.00	
2	HITTITE	1.908	0,37	??	23	-,		0.00	
3	LUVIAN	1.681	0,33	??	1		18.415,04	0,00	
4	LINGUISTIC	1.744	0,34	??	2.772		13.121,51	0,00	
5	SOCIOLINGUISTICS	1.015	0,20	??	67		10.626,01	0,00	
6	DIALECT	1.168	0,23	??	709			0,00	
7	SOCIOLINGUISTIC	639	0,12	??	129			0,00	
8	SPEAKER	1.074	0.21	??	8.830			0,00	
9 10	CTH	363	0.07	??	2			0,00	
11	ANATOLIAN	360	0.07	??	51		3.638,59		
12	SPEECH	859	0.17	??	9.811	0.02	3.523,92		
13	STUDY ED	1.239 591	0,24 0,12	??	31.553 3.192	0,03		0,00	
14	VARIATION	621	0,12	?? ??	3.895			0.00	
15	KUB	278	0,05	??	5.695 0		3.047.46		
16	MELCHERT	277	0.05	??	Ő			0.00	
17	SPEAK	784	0,15	??	10.487			0.00	
18	ENGLISH	1.090	0,21	??	27.147	0,02	2.944,53		
19	TEXT	711	0,14	??	8.527	0,02	2.853,34		
20	CODE	643	0.13	??	6.522			0.00	
21	ZA	302	0,06	??	133			0.00	
22	SOCIAL	1.302	0.25	??	45.508	0,04		0.00	
23	LDS	269	0,05	??	68			0,00	
24	ARZAWA	236	0,05	??	0		2.587,04	0,00	
25	KIZZUWATNA	235	0,05	??	0		2.576,07	0,00	
26	WORD	891	0,17	??	23.889	0,02	2.291,91	0,00	
27	FORM	1.085	0,21	??	38.429	0,03	2.274,57	0,00	
28	WA	315	0,06	??	623			0,00	
29	HATTUSA	206	0,04	??	1			0,00	
30	PUNJABI	234	0.05	??	72			0,00	
31 32	KBO	203	0.04	??	0		2.225.28		
33	STANDARD		0,15	??	16.510	0,01	2.217.08		
33	VARIETY	633	0.12	??	10.415			0.00	
35	SWITCH	459	0.09	??	3.803		2.151.09	0.00	
36	LABOV FISHMAN	239 207	0,05 0,04	?? ??	141 34			0,00	
37	WRITE	656	0.04	??	12.926	0,01		0.00	
38	ANATOLIA	218	0,13	??	12.920	0,01		0.00	
39	FEATURE	530	0,04	??	7.587			0.00	
40	CF	291	0,06	??	827			0.00	
41	GENITIVE	190	0,00	??	30			0.00	
42	SCRIBE	214	0.04	??	155			0.00	
13	DAKIOTAN	0.40	0.07				4 700 40	0,00	
		ource text	notes						
87 ent	trie Row 1			LANC	GUAGE				

Figure 8. Keywords for the Sociolinguistics subcorpus.

In this case, the list starts with the most key word in the corpus, according to its keyness⁶ (3). The example shows that the word LANGUAGE is the most key in the corpus;you can see that its frequency is 5.249 entries and its keyness is 30.930,87. Keyness is not related to frequency. If you compare the words SOCIOLINGUISTICS and DIALECT, you can notice that DIALECT is more frequent than SOCIOLINGUISTICS, but the keyness values are inverted.

Not all researchers use the Keyword tool for their researches. Nevertheless, it's a very useful tool for terminological works, the kind of work we're going to present in module 4.

Concordance

You can choose a word, from the Wordlist or Keywords screens, and ask for its concordance lines. The result, in a KWIC (keyword in context) screen, shows all the times the chosen word appears in the corpus, in example lines taken directly from the texts, and with the chosen word centralized in a different color. Let's observe the concordances for the word VALENCE (figure 9):

<u>File</u>	dit View Compute Settings Windows Help								
N	Concordance	1	Word #	Sent.#	Sent. Pos.	Para. #	ra. F Hea Hea	Sect. Set	File Date 4
1	66666666666666666666666666666666666666		88.791	3.061	53%	0 8	7%	0 87%	A CONSTRUCTIONAL APF 2011/nov/ 1
2	66666666666666666666666666666666666666		89.916	3.091	59%	0 8	8%	0 88%	A CONSTRUCTIONAL APF 2011/nov/ 1
3	either participant as highly individuated. Finally, the valence 3INITIATOR , patient4 in (10d) implies a		7.052	272	11%	0 4	7%	0 47%	A frame-based approach to 2013/mar/ 4
- 4	two participants from a specific perspective. The valence 3INITIATOR, location4 in (10a) takes the		6.819	265	9%	0 4		0 46%	A frame-based approach to 2013/mar/ 4
5	scene, characteristic of pattern (a). Conversely, the valence 3source, PATIENT4 in (10b) gives		6.851	266	8%	0 4			A frame-based approach to 2013/mar/ 4
6	four possible interpretations of the underspecified valence: 3source, location4, 3source, patient4,		6.207	234	41%	0 4	2%		A frame-based approach to 2013/mar/ 4
7	in the overall interpretation of each alternation. The valence 3source, location4, instantiated in (10c), is		6.751	262	10%	0 4			A frame-based approach to 2013/mar/ 4
8	indirect causation, respectively. This suggests the valence 3start, patient4. Note that the verb		9.728	388	71%	0 6			A frame-based approach tc 2013/mar/ 6
9	are associated with the single, underspecified valence 3start, end4 (enclosed in a rectangle)		6.286	236	76%	0 4			A frame-based approach tc 2013/mar/ 4
10) associated with the lexically specified predicate valence. 5. Conclusions This paper is a study of the		11.323	444	100%	0 7			A frame-based approach to 2013/mar/ 7
11	81 DERIVATIONAL SUFFIXES NOT AFFECTING VALENCE		328	23	100%	0			VALENCE CHANGE AND (2011/nov/ 4
12	may be explained in terms of differential emotional valence across the languages mastered by a given		8.616	371	60%	0 6			Clinical Neurolinguistics.txt 2014/abr/l 6
13	Transitive, Intransitive, Intransitive, Transitive, valence active subjects often passive active subjects		30.003	1.169	22%	0 4			Constructional Morphology. 2011/nov/. 7
14	combinations are complex. The impact of each valence-affecting suffix has been discussed in		23.600	1.613	29%	0 8			VALENCE CHANGE AND (2011/nov/ 1
15	their interaction. Below I examine a sampling of valence-affecting suffix combinations to see if the		23.622	1.614	36%	8 0			VALENCE CHANGE AND (2011/nov/ 1
16	be emphasized that version is not simply part of a valence alternation that derives indirect objects norm		51.654	1.964	68%	0.6			Constructional Morphology. 2011/nov/: 1
17	pre x participates in a number of version-like and valence alternation constructions, suggesting a		10.916	398	62%	0 1			Constructional Morphology. 2011/nov/: 2
19	distinctions on the verbs corresponding to traditional valence alternations. Instead, he suggests that		67.802	2.636	98%	0 9			Constructional Morphology. 2011/nov/: 1
20	next section examines some more cases of voice/valence alternations that exhibit constructional		67.244	2.621	31%	0 8			Constructional Morphology. 2011/nov/: 1
20	for representing argument structure constraints, valence alternations, and the relation between		12.708	476	57%	0 1			Constructional Morphology. 2011/nov/. 2
21	as a family of related constructions that involve valence alternations. The chapter extends the		28.218	1.088	95%	0 3			Constructional Morphology. 2011/nov/. 7
23	and indirect object coding, as well as voice and valence alternations. I will analyze the		7.124	254	95%	0			Constructional Morphology. 2011/nov/. 1
24	framework by adding a mechanism to handle valence alternations. A representation for relations		28.232	1.089	93%	0 3			Constructional Morphology. 2011/nov/: 7
24	than one argument is added to a verb?s original valence and after examining the forms , $\Box k \Box k$ and ,		17.103	1.186	12%	0 6			VALENCE CHANGE AND (2011/nov/ 1
26	the "role invariance" (all variants have the same valence) and the "role diversity" (each variant has a		6.640	259	69%	04			A frame-based approach tr 2013/mar/ 4
27	verb sumixes and ten amerent sumixes when aneot valence and/or grammatical relations. Two sumixes		1.102	116	85%	0			VALENCE CHANGE AND (2011/nov/ 1.
28	, a single verb is usually associated with only one valence and word-order pattern. This suggests that		26.985	1.031	85%	0 3			Constructional Morphology. 2011/nov/. 6
29	e, vereien is lightly soupled with the frame system, fallence and voice alternation constructions, and		10.158	377	33%	0 1			Constructional Morphology. 2011/nov/. 2
30	to mention my ice-climbing mentor, partners Nicolas Valence and Dave Harris, also Jonah Flower and		2.542	96	69% 86%	0 5			The syntax of verbal inflect 2012/mai/ 2 A frame-based approach to 2013/mar/ 4
31	event type imposed by the 3INITIATOR, PATIENT4 valence and the lexical meaning of the predicate. motivation constructs. Her findings suggested that valence and self-efficacy are critical in predicting		7.432 9.639	286 389	20%	0 1			A frame-based approach to 2013/mar/ 4 IDENTITY AND SECOND I 2013/mar/ 2
32			9.639	389 59	20% 94%	0			
33	. I have examined five suffixes which decrease valence and five which increase valence, as well as				94% 36%	0 9			Constructional Morphology. 2011/nov/: 3 VALENCE CHANGE AND (2011/nov/ 1
34			27.304 15.319	1.841	13%	0 5			VALENCE CHANGE AND (2011/nov/ 1 VALENCE CHANGE AND (2011/nov/ 1
35	noun phrases can also be added to a verb?s valence and marked by the applicative, but a ideally suited to handle. 3.2.2 Conjugation Classes, Valence, and Case-Marking Traditionally, Georgian		29.781	1.159	29%	0 3			Constructional Morphology. 2011/nov/: 7
36	often, the elements in this slot indicate changes in valence, and case-marking Traditionally, Georgian		43.314	1.671	29%	0.5			Constructional Morphology. 2011/nov/. 1 Constructional Morphology. 2011/nov/. 1
37	onen, the elements in this slot indicate changes in valence and grammatical voice. However, the		43.314	1.0/1	200/	0 5		0 01%	Towarda Tavt Knowladge 5 2012/mai/ 2
<									>
-	acce carcates plot patterns clusters timeline filenames sourcetext notes								
347 ent	ries Row 1 64 category 2 666664 head adj valence 2 64 subj hNP:5 i comp								
			DIGE		~ .	· · ·	. ~		

Figure 9. Concordance lines for the word VALENCE in the Linguistics Corpus.

As you can see, the chosen word (VALENCE) is centralized and its color is different (blue). The first word to the right of VALENCE (in red) is classified alphabetically. In this example, we can find 347 lines of concordances (4). In WordSmith Tools, if you click twice in the line you want to analyze, the text is going to be open in other tab, with that word highlighted. Let's analyze the line 16 of picture 4 by double clicking it (figure 10):

⁶ Keyness is how key the word is. To get to this number, the program compares your study corpus with a reference corpus (which you must get previously) and statiscally calculates the probabilities of usage in each one.

<u>File Edit View Compute Settings Windows H</u> elp
with the pre-stem slot there are exceptions on both formal and functional sides of the
relationship.
It should also be emphasized that version is not simply part of a valence alternation
that derives indirect objects from obliques. The version constructions themselves impose semantic and pragmatic constraints that should not be expected from a lexical rule-like
operation. Verbs with version do not always have a corresponding versionless verb (at least
synchronically), and the precise meaning of a verb with version is not always predictable
based on a versionless variant.
Michaelis and Ruppenhofer (2001) come to a similar conclusion in the examination
of the German applicative with the preffix be At
rst glance, the German be-pattern seems
to be a simple argument promotion device parallel to the locative alternation in English,
e.g. schmieren `to smear X on Y' beschmieren `to smear Y with X'. This pattern accepts
both transitive verbs of transfer and bivalent intransitive verbs of locomotion, e.g. wandern
`wander', in each case promoting what would otherwise be an oblique to a core argument
(usually direct object). Also, as the English locative alternation, the German applicative is
usually described as applying to a limited semantic class of verbs. However, Michaelis and
Ruppenhofer's extensive examination of corpus data involving this pattern reveals that the
applicative pattern has its own unique properties that cannot be derived via lexical rules from non-applicative verb entries. First, in some instances the applicative pattern assigns
grammatical relations that would not be expected from the input verbs. So in (154), the
`input' lexical entry `cook' does not license a goal argument, yet the preffixed verb assigns
core argument status to a semantic goal or bene
ciary.
(154) Aber wenn ich es mir recht 🛛 uberlege habe ich mich von meinen Kollegen [] auch
mal eifrig Ka
e bekochen lassen []
But, if I remember correctly, at times I also let myself be be-cooked with co
ee by
my colleagues.' (Michaelis and Ruppenhofer 2001:3)
Second, the corpus data also suggests that the applicative pattern is less selective 157
about what verbs serve as input, questioning the assumption of constrained productivity.
Third, the pattern seems to accept even nouns as input, as in (155).
(155) As mag ja lustig sein, zwei hartgekochte Eier wie Clownski opfe mit angekeimten
Sojabohnen zu behaaren und sie auf Gurkenscheiben zu stellen, ihnen mit zwei
Tomatenstreifen Munder zu verpassen und Auglein aus Sojasprossen einzudr ucken.
Ok, it might be funny to be-hair two hard-boiled eggs like clown's heads with
germinating soy beans, to stand them up on cucumber slices, to give them mouths
from tomato strips, and to impress soy shoots on them as little eyes.' (Michaelis
and Ruppenhofer 2001:5)
Instead of looking for the right way to constrain the potential inputs to the applicative concordance collocates plot patterns clusters timeline filenames source text notes
0% 25 found rsion is not simply part of a valence alternation that deriv

Figure 10. Text tab for the word VALENCE; line 16 (valence alternation) chosen.

As you can see, we have the chosen example (valence alternation) highlighted in the second paragraph of the text. The text can be copied using the control C command.

Generally, the concordance lines are the starting point of your research. It's from here that you're going to identify linguistic phenomena that are pertinent for your studies.

Collocates

According to McGlashan (2013), collocates are [...] "A co-occurrence relationship between words or phrases. Words aresaid to collocate with one another if one is more likely to occur in the presence of the other than elsewhere." If we analyze the collocates tab in the VALENCE concordance screen, we get the following results (Figure 11):

VALENCE valence 0,000 38 353 4 4 1 2 1 345 1 2 1 2 THE valence 0,000 24 204 137 67 15 25 19 15 63 3 20 16 14 14 3 OF valence 0,000 24 27 36 41 6 3 5 7 16 19 30 9 5 4 5 4 AND valence 0,000 24 77 36 41 6 3 5 3 19 16 6 8 4 7 5 IN valence 0,000 22 54 33 21 3 4 6 12 8 7 7 4 3 1 9 IS valence 0,000 13 38 19 19 1 4 7 7 1 11 5 2 3 4 5 2<	0															Cond	ord		
VALENCE valence 0.00 38 353 4 4 1 2 1 345 1 2 1 1 THE valence 0.000 24 204 137 67 15 25 19 15 63 3 20 16 14 14 3 OF valence 0.000 24 277 36 41 63 5 3 19 16 6 8 4 7 7 7 7 6 4 valence 0.000 22 54 33 21 3 4 6 12 8 7 7 7 6 2 7 7 6 2 1 11 7 6 2 1 4 3 11 11 15 4 8 6 5 2 1 4 3 11 15 2 8 1 4 2 5 3 4 4 3 11 11 11 11 11	<u>F</u> ile <u>E</u>	dit <u>V</u> iew <u>C</u> o	ompute	<u>S</u> ettings	<u>W</u> indows	: <u>H</u> elp)												
VALELOL Valence 0.00 24 24 24 24 14 14 15 25 19 15 63 3 20 16 14 14 3 OF valence 0.000 24 204 137 67 15 25 19 15 63 3 20 16 14 14 4 AND valence 0.000 24 277 36 41 6 3 5 3 19 16 6 8 4 7 5 IN valence 0.000 16 67 33 21 3 4 6 12 8 7 7 4 3 6 A valence 0.000 12 39 11 5 4 7 1 11 5 2 1 4 5 1 8 5 2 1 4 5 2 1 4 1 11 15 2 1 11 14 1 1	N	W	/ord	With	Relation	Texts	Total	Total Left	Total Right	L5	L4	L3	L2	L1 Centre	R1	R2	R3	R4	R5
3 OF valence 0,000 24 107 <td< td=""><td>1</td><td>VALENO</td><td>CE</td><td>valence</td><td>0,000</td><td>38</td><td>353</td><td>4</td><td>4</td><td>1</td><td>2</td><td></td><td>1</td><td>345</td><td></td><td>1</td><td></td><td>2</td><td>1</td></td<>	1	VALENO	CE	valence	0,000	38	353	4	4	1	2		1	345		1		2	1
4 AND valence 0,000 22 77 36 41 6 3 5 3 19 16 6 8 4 7 5 IN valence 0,000 16 67 36 31 9 4 4 9 10 6 5 7 7 6 6 A valence 0,000 12 39 12 3 4 6 12 8 7 7 4 3 1 7 THAT valence 0,000 12 39 19 20 4 8 1 4 2 5 3 4 5 3 11 11 5 4 8 6 5 2 1 4 3 11 15 2 3 4 5 3 4 5 3 4 5 3 4 5 3 4 5 3 4 5 3 4 5 3 4 5 3 4 5<	2	TH	HE	valence	0,000	24	204	137	67	15	25	19	15	63	3	20	16	14	14
5 IN valence 0,000 24 i 0 <	3	(DF	valence	0,000	25	105	52	53	5	5	7	16	19	30	9	5	4	5
6 A valence 0.000 12 54 33 21 3 4 4 6 12 8 7 7 4 3 7 THAT valence 0.000 14 40 23 17 6 5 4 7 1 7 6 2 2 8 TO valence 0.000 12 39 19 20 4 8 1 4 2 5 3 4 5 3 9 USR valence 0.000 13 38 19 19 1 4 7 7 1 11 5 2 13 BY valence 0.000 6 31 8 23 2 2 1 11 1 3 4 4 5 3 4 4 5 3 4 4 5 1 1 1 1 1 1 1 1 3 5 3 2 1 1 1 1	4	A	ND	valence	0,000	24	77	36	41	6	3	5	3	19	16	6	8	4	7
7 THAT valence 0,000 14 40 23 17 6 5 4 7 1 7 6 2 2 8 TO valence 0,000 12 39 19 20 4 8 1 4 2 5 3 4 5 3 1 5 2 1 4 3 1 1 1 5 2 1 4 3 1 1 1 5 2 1 4 3 1 1 1 5 2 1 4 3 1 1 1 5 2 1 4 3 1 1 1 5 2 1 4 3 1 1 1 5 2 1 4 3 1 1 1 3 4 4 2 2 1 1 1 3 4 4 2 1 1 1 1 3 3 3 3 3 3 3 3	5		IN	valence	0,000	16	67	36	31	9	4	4	9	10	6	5	7	7	6
7 THAT valence 0,000 14 40 23 17 6 5 4 7 1 7 6 2 2 8 TO valence 0,000 12 39 19 20 4 8 1 4 2 5 3 4 5 3 9 IS valence 0,000 13 38 19 19 1 4 7 7 1 11 15 3 4 4 7 13 As valence 0,000 1 31 12 19 2 1 6 3 13 4 4 2 13 BY valence 0,000 8 24 9 15 2 3 2 1 1 14 4 4 4 4 4 4 4 4 4 4 3 2 1 13 3 3 3 3 3 3 3 3 3 3 3 3 <td>6</td> <td></td> <td>Α</td> <td>valence</td> <td>0,000</td> <td>22</td> <td>54</td> <td>33</td> <td>21</td> <td>3</td> <td>4</td> <td>6</td> <td>12</td> <td>8</td> <td></td> <td>7</td> <td>7</td> <td>4</td> <td>3</td>	6		Α	valence	0,000	22	54	33	21	3	4	6	12	8		7	7	4	3
3 IS valence 0,000 12 39 19 20 4 8 1 4 2 5 3 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 8 5 1 1 11 15 valence 0,000 13 38 19 19 2 1 6 3 13 4 2 3 1 11 11 13 4 4 5 1 8 5 1 11 13 4 4 5 1 1 11 13 4 4 5 1 13 14 5 1 11 13 4 4 5 13 4 4 1 11 13 12 1 13 13 14 1 13 13 13 13 13 13 13 13 13	7	TH/	AT	valence	0,000	14	40	23	17	6	5	4	7	1		7	6	2	2
10 VERB valence 0,000 12 38 19 19 1 4 7 7 1 11 5 2 11 AS valence 0,000 13 38 19 19 6 4 5 4 5 1 8 5 1 14 2 1 11 15 2 1 11 11 13 4 4 2 13 4 4 2 13 4 4 2 13 4 4 4 5 1 11 14 4 5 1 8 5 13 4 4 2 1 13 4 4 4 4 5 1 13 4 4 4 5 1 13 4 4 4 4 3 11 14 4 4 1 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	8	٦	го	valence	0,000	9	39	28	11	5	4	8	6	5	2	1	4	3	1
11 AS valence 0,000 13 13 19 19 6 4 7 1 11 18 5 12 SUFFIXES valence 0,000 1 31 12 19 2 1 6 3 13 4 2 13 BY valence 0,000 6 31 8 23 2 2 1 2 1 11 14 4 4 5 14 7 valence 0,000 2 25 16 9 1 1 14 4 4 5 15 ARE valence 0,000 7 22 12 10 3 4 3 2 1 3	9		IS	valence	0,000	12	39	19	20	4	8	1	4	2	5	3	4	5	3
12 SUFFIXES valence 0,000 1 31 12 19 2 1 6 3 13 4 4 13 BY valence 0,000 6 31 8 23 2 2 1 2 1 11 1 3 4 4 14 7 valence 0,000 2 25 16 9 1 1 14 4 5 2 16 WITH valence 0,000 7 22 12 10 3 4 3 2 1 3 2 3 2 </td <td>10</td> <td>VEF</td> <td>RB</td> <td>valence</td> <td>0,000</td> <td>8</td> <td>38</td> <td>19</td> <td>19</td> <td>1</td> <td></td> <td>4</td> <td>7</td> <td>7</td> <td></td> <td>1</td> <td>11</td> <td>5</td> <td>2</td>	10	VEF	RB	valence	0,000	8	38	19	19	1		4	7	7		1	11	5	2
13 BY valence 0,000 6 31 12 13 2 2 1 2 1 1 1 1 3 4 4 14 7 valence 0,000 2 25 16 9 1 1 14 4 4 5 2 15 ARE valence 0,000 7 22 12 10 3 4 3 2 1 3		ł	AS	valence	0,000	13	38	19	19	6	4	5	4		5	1	8	5	
14 7 valence 0,000 2 25 16 9 1 1 14		SUFFIXE	ES	valence	0,000	1	31	12	19	2	1	6	3			13	4		2
15 ARE valence 0,000 8 24 9 15 2 3 4 1 5 2 16 WITH valence 0,000 7 22 12 10 3 4 3 2 1 3 1 2 1 2 3 3 1 2 1 1 1 2 1 2 1 3 3 3 3 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	13	E	ΒY	valence	0,000	6	31	8	23	2	2	1	2	1	11	1	3	4	4
16 WITH valence 0,000 7 22 12 10 3 4 3 2 1 3 1 2 1 3 5 3 2 1 1 3 5 3 2 1 1 1 3 5 3 2 1	14		7	valence	0,000	2	25	16	9		1	1	14					4	5
17 CHAPTER valence 0,000 2 21 18 3 2 1 13 2 14 1 13 12 14 11 14 12 13 13 2 16 15 15 13 13 14 11 14 12 14 12 <th12< th=""> 13 14</th12<>	15	AF	RE	valence	0,000	8	24	9	15	2	3	2		2	3	4	1	5	2
18 OR valence 0,000 15 21 16 15 2 1 16 2 5 14 1 1 2 1 1 1 3 8 3 1 2 1 19 S valence 0,000 6 21 8 13 2 4 1 1 3 5 3 2 1 20 VERBS valence 0,000 6 20 9 11 4 2 3 2 6 3 21 2 valence 0,000 6 20 9 11 4 2 3 2 6 3 22 ONE valence 0,000 6 20 15 5 2 6 2 4 1 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 1		W	TH	valence	0,000	7	22	12	10	3	4	3		2	1	3		3	3
19 Si valence 0,000 3 21 21 0 2 5 14 20 VERBS valence 0,000 6 21 8 13 2 4 1 1 3 5 3 2 21 2 valence 0,000 6 20 9 11 4 2 3 2 6 3 22 ONE valence 0,000 6 20 9 11 4 2 3 2 6 3 22 ONE valence 0,000 6 20 15 5 2 6 2 4 1 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 3 3 2 16 1 1 1 14 2 1 1 1 14 2 1 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1		CHAPTE	ER	valence	0,000	2	21	18	3	2	1	13	2			1			2
20 VERBS valence 0,000 6 21 2 1 1 1 1 3 5 3 2 21 2 valence 0,000 6 20 9 11 4 2 3 2 6 3 22 ONE valence 0,000 6 20 9 11 4 2 3 2 6 3 22 ONE valence 0,000 3 20 3 17 3 9 3 2 3 23 3 valence 0,000 6 20 15 5 2 6 2 4 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 1 1 1 4 2 1 1 1 1 1 1 1 1 1 1 1 1 <		C	DR	valence	0,000	15		6	15	2	1			3	8	3	1	2	1
21 2 valence 0,000 6 20 9 11 4 2 3 2 6 3 22 ONE valence 0,000 6 20 9 11 4 2 3 2 6 3 22 ONE valence 0,000 6 20 15 5 2 6 2 4 1 2 1 1 1 1 4 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1			S	valence	0,000	3	21	21	0		2		5	14					
22 ONE valence 0,000 3 20 3 11 4 2 3 2 3 3 23 3 valence 0,000 6 20 15 5 2 6 2 4 1 2 1	20	VERE	BS	valence	0,000	6	21	8	13	2	4	1		1		3	5	3	2
23 3 valence 0,000 6 20 15 5 2 6 2 4 1 2 1			2	valence	0,000	6	20	9	11		4	2		3	2			6	3
24 SPECIFIERS valence 0,000 1 19 19 19 25 CHANGE valence 0,000 1 19 19 19 26 USING valence 0,000 6 18 11 7 4 7 4 2 1 26 USING valence 0,000 3 18 2 16 1 1 14 2 1 26 USING valence 0,000 4 17 1 16 1 1 14 2 1 27 PATTERNS valence 0,000 1 16 16 0 2 14 1 29 SUFFIX valence 0,000 5 16 11 5 2 2 3 4 1 2 2 30 WHICH valence 0,000 2 15 0 15 15 15 14 1 32 ANALYSIS valence 0,000 1 15 15		0	NE	valence	0,000	3		3	17					3		9		2	3
25 CHANGE valence 0,000 6 18 11 7 4 7 4 2 1 26 USING valence 0,000 3 18 2 16 1 1 14 2 1 26 USING valence 0,000 3 18 2 16 1 1 14 2 1 27 PATTERNS valence 0,000 4 17 1 16 1 15 1 15 1 15 1 2 2 3 4 1 2 2 3 4 1 2 2 3 4 1 2 2 3 4 1 2 2 3 4 1 2 2 3 4 1 2 2 3 4 1 2 2 3 4 1 2 2 3 4 1 2 2 3 4 1 2 2 3 15 15 15 14			3	valence	0,000	6	20	15	5	2	6	2	4	1			2	1	2
26 USING valence 0,000 3 18 2 16 1 1 14 2 27 PATTERNS valence 0,000 4 17 1 16 1 1 14 2 28 DETERMININ valence 0,000 4 17 1 16 1 1 14 2 29 SUFFIX valence 0,000 1 16 10 6 1 2 7 4 2 30 WHICH valence 0,000 5 16 11 5 2 2 3 4 1 2 2 30 WHICH valence 0,000 5 16 11 5 2 2 3 4 1 2 2 31 INCREASING valence 0,000 2 15 0 15 15 14 1 32 ANALYSIS valence 0,000 1 15 0 15 14 1 33 </td <td></td> <td>SPECIFIE</td> <td>RS</td> <td>valence</td> <td>0,000</td> <td>1</td> <td>19</td> <td>0</td> <td>19</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>19</td> <td></td> <td></td> <td></td> <td></td>		SPECIFIE	RS	valence	0,000	1	19	0	19						19				
27 PATTERNS valence 0,000 4 17 1 16 1 15 1 28 DETERMININ valence 0,000 1 16 16 0 2 14 15 1 29 SUFFIX valence 0,000 1 16 10 6 1 2 7 4 2 30 WHICH valence 0,000 5 16 11 5 2 2 3 4 1 2 2 30 WHICH valence 0,000 5 16 11 5 2 2 3 4 1 2 2 31 INCREASING valence 0,000 2 15 0 15 15 14 1 32 ANALYSIS valence 0,000 1 15 0 15 14 1 33 SENTIMENT valence 0,000 1 14 1 13 1 9 2 1 1				valence	0,000							4			4	2			1
28 DETERMININ valence 0,000 1 16 16 0 2 14 29 SUFFIX valence 0,000 1 16 10 6 1 2 7 4 2 30 WHICH valence 0,000 5 16 11 5 2 2 3 4 1 2 2 30 WHICH valence 0,000 5 16 11 5 2 2 3 4 1 2 2 31 INCREASING valence 0,000 2 15 0 15 15 14 1 32 ANALYSIS valence 0,000 1 15 0 15 14 1 33 SENTIMENT valence 0,000 1 14 1 13 1 9 2 1 1 34 SPECIFIER valence 0,000 3 13 0 13 13 13 13 36 BE				valence	0,000	3							1	1					
29 SUFFIX valence 0,000 1 16 10 6 1 2 7 4 2 30 WHICH valence 0,000 5 16 11 5 2 2 3 4 1 2 2 31 INCREASING valence 0,000 5 16 11 5 2 2 3 4 1 2 2 31 INCREASING valence 0,000 2 15 0 15 15 14 1 32 ANALYSIS valence 0,000 1 15 0 15 14 1 33 SENTIMENT valence 0,000 1 15 0 15 14 1 34 SPECIFIER valence 0,000 1 14 1 13 1 9 2 1 1 36 BE valence 0,000 5 13 6 7 1 2 3 5 2					,					1					15		1		
30 WHICH valence 0,000 5 16 11 5 2 2 3 4 1 2 2 31 INCREASING valence 0,000 2 15 0 15 15 15 15 32 ANALYSIS valence 0,000 1 15 0 15 14 1 33 SENTIMENT valence 0,000 1 15 0 15 14 1 34 SPECIFIER valence 0,000 1 15 0 15 14 1 34 SPECIFIER valence 0,000 1 14 1 13 1 9 2 1 1 36 BE valence 0,000 5 13 6 7 1 2 3 5 2													2	14					
31 INCREASING valence 0,000 2 15 0 15 15 15 32 ANALYSIS valence 0,000 1 15 0 15 14 1 33 SENTIMENT valence 0,000 1 15 0 15 14 1 34 SPECIFIER valence 0,000 1 15 0 15 14 1 34 SPECIFIER valence 0,000 1 14 1 31 1 9 2 1 1 35 CHANGING valence 0,000 3 13 0 13 13 36 BE valence 0,000 5 13 6 7 1 2 3 5 2		SUFF	IX	valence	0,000		16									4			
32 ANALYSIS valence 0,000 1 15 14 1 33 SENTIMENT valence 0,000 1 15 14 1 34 SPECIFIER valence 0,000 1 14 1 13 1 9 2 1 1 36 BE valence 0,000 5 13 6 7 1 2 3 5 2					'	-			-	2	2	3	4				2		2
33 SENTIMENT valence 0,000 1 15 0 15 14 1 34 SPECIFIER valence 0,000 1 14 1 13 1 9 2 1 1 35 CHANGING valence 0,000 3 13 0 13 13 36 BE valence 0,000 5 13 6 7 1 2 3 5 2					'	_		-							15				
34 SPECIFIER valence 0,000 1 14 1 13 1 9 2 1 1 36 CHANGING valence 0,000 3 13 0 13 13 13 36 BE valence 0,000 5 13 6 7 1 2 3 5 2					-,			-									14		1
35 CHANGING valence 0,000 3 13 13 36 BE valence 0,000 5 13 6 7 1 2 3 5 2					'			-								14			
³⁶ BE valence 0,000 5 13 6 7 1 2 3 5 2					,					1							2	1	1
								-							13				
³⁷ ROOT valence 0,000 4 13 9 4 1 1 7 4						5				1		3				5	2		
		ROC	DT	valence	0,000	4	13	9	4		1		1	7				4	
concordance collocates plot patterns clusters timeline filenames source text notes		ance collocates	plot	patterns c	lusters tin	neline	filename	s source	text notes	^	<u>^</u>				1		^		^

Figure 11. Collocates for the word VALENCE·

As you can realize, the word VALENCE is in the first position, and the total value (345) is in the centre column (in red). Let's pretend we want to analyze the relationship between the words VALENCE and SUFFIXES (line 12); we can see that SUFFIXES appears 13 times (in red) two positions right from VALENCE. If you want to study more this relationship, just click twice in the number 13 and you get back the concordance tab, with these two words chosen (figure 12).

MÓDULO 1

	Concord						_ 5
[ile]	dit <u>V</u> iew <u>Compute</u> <u>Settings</u> <u>Windows</u> <u>Help</u>						
N	Concordance	§ 1 Word #	Sent. #	Sent. Pos.	Para. # Para.	Sect. I Hez Hez Sec	r File Date
1	100 Table 11. Attested Combinations of Valence-Changing Suffixes	520	63	83%	0 2%	0 2%	VALENCE CHANGE AND (2011/nov/
2	39 3.6 Summary of Valence Reducing Suffixes.	287	18	75%	0 1%	0 1%	VALENCE CHANGE AND (2011/nov/
3	in all of the morphological verb-forming processes. Valence-changing suffixes may combine with each	24.370	1.662	13%	0 85%		VALENCE CHANGE AND (2011/nov/
- 4	, □m). 100 101 Table 11. Attested Combinations of Valence-Changing Suffixes Derivation Position 1	23.489	1.607	17%	0 82%		VALENCE CHANGE AND (2011/nov/
5	v s-0? -vis other consitutients. 4.4 Summary of Valence Increasing Suffixes I here summarize the	19.497	1.336	31%	0 68%		VALENCE CHANGE AND (2011/nov/
6	73 4.4 Summary of Valence Increasing Suffixes .	320	22	75%	0 1%		VALENCE CHANGE AND (2011/nov/
7	10 Table 3. Valence Reducing Suffixes.	456	47	50%	0 2%		VALENCE CHANGE AND (2011/nov/
8	Verb Roots) 69 Table 6. Valence Increasing Suffixes .	482	53	50%	0 2%		VALENCE CHANGE AND (2011/nov/
9	the applicative suffix ,do. 3.6 Summary of Valence Reducing Suffixes The most important	10.432	752	38%	0 36%		VALENCE CHANGE AND (2011/nov/
10	most important information about Mbonge?s five valence-reducing suffixes is summarized in Table 3.	10.442	753	33%	0 36%		VALENCE CHANGE AND (2011/nov/
11	reasons and with different results. Table 3. Valence Reducing Suffixes Suffix Description	10.468	757	2%	0 37%		VALENCE CHANGE AND (2011/nov/
12	reasons and with different results. 82 Table 6. Valence Increasing Suffixes Suffix Description	19.536	1.341	7%	0 68%		VALENCE CHANGE AND (2011/nov/
13	11 shows all the attested suffix combinations of the valence-changing suffixes. Table 10. Attested	23.294	1.588	92%	0 81%		VALENCE CHANGE AND (2011/nov/
14	282 5.3.1.3. Differences in the valence patterns .	1.959	160	88%	0 2%		Criteria for the Validation.tx 2013/mai/
15	2012) 130 Also, a second table illustrates the valence patterns of the LU, i.e. its combinatorial	40.322	1.796	41%	0 36%		Criteria for the Validation.tx 2013/mai/.
16	the FE is further enriched with the specification of valence patterns, the semantic and syntactic	4.118	182	48%	0 50%		Process-oriented terminolo 2012/mai/
17	realizations of the different FEs, as well as their valence patterns or mappings between semantic and	3.506	161	77%	0 43%		Process-oriented terminolo 2012/mai/
18	issues? RQ1c Can we automatically determine the valence (positive, negative) of these semantic	4.744	187	57%	0 21%		Semantic Network Analysis 2011/nov/
19	(who is acting and who is acted upon) and sign (or valence, polarity: is the relation positive or	33.053	1.296	77%	1 97%		Semantic Network Analysis 2011/nov/
20	and a constructional description of the TAM and valence patterns resolve the apparent	29.376	1.139	70%	0 39%		Constructional Morphology. 2011/nov/.
21	constituents are NP + VP + ADVP. Table 6. Valence patterns of TIDE Table 7. Valence patterns	4.189	186	29%	0 51%		Process-oriented terminolo 2012/mai/
22	, resolve1 and suprir1. 5.3.1.3. Differences in the valence patterns About 6% of the total number of	83.552	3.582	10%	0 75%		Criteria for the Validation.tx 2013/mai/
23	of determinar2. The Portuguese verb admits three valence patterns: JUDGE (Sub. NP)	83.838	3.593	70%	0 75%		Criteria for the Validation.tx 2013/mai/.
24	structures of the verbs revealed that the verbs' valence patterns are different. For instance,	83.602		95%	0 75%		Criteria for the Validation.tx 2013/mai/.
25	130 Figure 21. Lexical entry report of argue: valence patterns (FrameNet 2012) 131	2.853	297	70%	0 3%		Criteria for the Validation.tx 2013/mai/.
	ADVP. Table 6. Valence patterns of TIDE Table 7. Valence patterns of MAREA (1a) TIDE to occur	4.195	187	13%	0 51%		Process-oriented terminolo 2012/mai/
27	, and prepositions as well as to identify their valence patterns (the ways in which the semantic	39.017	1.726	74%	0 35%		Criteria for the Validation.tx 2013/mai/.
28	occur in the contexts of determinar2. Among the valence patterns that the term require1 admits the	83.969	3.613	18%	0 75%		Criteria for the Validation.t/ 2013/mai/
29	njn ?kd qm?. Each suffix indicates an increase in valence, resulting in a net addition of two 104	24.184	1.651	47%	0 84%		VALENCE CHANGE AND (2011/nov/
30	in section 3.2.1.3. For this reason, the sense and valence rich descriptions provided by FrameNet and	40.783	1.815	23%	0 36%		Criteria for the Validation.tv 2013/mai/
31	a ectedness marking, to a set of voice and valence-related alternations. However, all of these	64.565	2.506	98%	0 86%		Constructional Morphology. 2011/nov/.
32	NP the bottom of the barrel is empty because the valence requirement has been satis ed. 4.1.	57.438		89%	0 56%		A CONSTRUCTIONAL APF 2011/nov/
33	Crime Safety Public Order & Safety Crime Justice Valence Rule of Law Economic Growth Economic	56.990	2.231	51%	3 6%		Semantic Network Analysis 2011/nov/
34	Person Definite Object) (Agreement= Verb root -Valence -Scenario -Reflexive/ -Negation -N-Object	6.224	374	44%	0 16%		The Morphology of Modern 2011/nov/
35	Person Indefinite Object) (Agreement= Verb root -Valence -Scenario -Reflexive/ -Negation -N-Object	6.252	374	73%	0 16%		The Morphology of Modern 2011/nov/
36	. V. VI. VII. VIII. IX. X. XI. Agreement= Verb root -Valence -Scenario -ReflexValence -Negation	6.628	404	71%	0 17%		The Morphology of Modern 2011/nov/
< 37	Demon and Casand Demon) Agrooments Vieth rest Violance Casanaria Magatian N Object Cantral	6 202	274	220/	0 100	0 100/	The Membelony of Medern 2011/nev/
concord							
347 entr	les Row 33 Order & Safety Crime Justice Valence Rule of Law Economic G						

Figure 12. Concordance lines with the words VALENCE and SUFFIXES

Just with this VALENCE example, we can study hundreds of combinations. For example, if you analyze picture 6, you can notice what are the prepositions that follow VALENCE one position to the right: of (30 times), in (6 times), by (11 times), etc. From this analysis, you could postulate that the most common preposition that follows the word VALENCE is of. A very simple postulation, but based on facts.

Clusters/N-Grams

According to Scott (2012), clusters are:

[...] words which are found repeatedly together in each others' company, in sequence. They represent a tighter relationship than collocates, more like multi-word units or groups or phrases. (I call them clusters because groups and phrases already have uses in grammar and because simply being found together in software doesn't guarantee they are true multi-word units.) Biber calls them "lexical bundles".

Language is phrasal and textual. It is not helpful to see it as a matter of selecting a word to fill a grammatical "slot" as implied by structural theories. Words keep company: the extreme example is idiom where they're bound tightly to each other, but all words have a tendency to cluster together with some others. These clustering relations may involve colligation (e.g. the relationship between depend and on), collocation, and semantic prosody (the tendency for cause to come with negative effects such as accident, trouble, etc.).

Following the same example above, let's see the clusters (from 3 to 5 words, with a minimum of 5 examples) involving VALENCE (Figure 13):

Basic principles of Corpus Linguistics

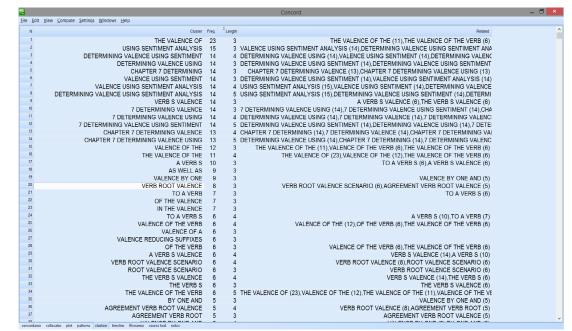


Figure 13. Cluster tab for the word VALENCE

In the example 20, we have the cluster verb⁷ root valence, with 8 examples. If you want to analyze this cluster, just go back to the collocates tab (take a look at line 37 of figure 11) and choose root. The result is the following (figure 14):

N Concordance	1 Word t Ser Ser Par Par Her Her Ser	Sci File Date	%
1 . IV. V. VI. VII. VIII. IX. X. XI. Agreement= Verb root -Valence -Scenario -ReflexValence -Negation N-Object	6.62 4(7' 0 1; 0	17 The Morpl 2011/nov/	37%
² Person and Second Person) Agreement= Verb root -Valence -Scenario -Negation -N-Object -Central -Tense		16 The Morpl 2011/nov/	
3 (Third Person Definite Object) (Agreement= Verb root -Valence -Scenario -Reflexive/ -Negation -N-Object		16 The Morpl 2011/nov/	
4 Positions I. II. III. IV. V. VI. VII. VII. IX. X. Verb root -Valence -Scenario Animacy -Central Participant/		18 The Morpl 2011/nov/	
5 Verb Structure Animate Subject (Agreement Verb root -Valence -Negation -N-Object -Person -Tense Clitic)=		1! The Morpl 2011/nov/	
Marker Agreement Inanimate Subject Verb root -Valence -Negation -Tense Specifier 27 Animate		1: The Morpl 2011/nov/	
7 (Third Person Indefinite Object) (Agreement= Verb root -Valence -Scenario -Reflexive/ -Negation -N-Object		16 The Morpl 2011/nov/	
entities; and 4) display no significant differences in their valence patterns. Equivalents were considered partial		76 Criteria fo 2013/mai	
9 [Reciprocality]. Figure 21. Lexical entry report of argue: valence patterns (FrameNet 2012) 132 The Using and	40.4 1.24 0 36 0	36 Criteria fo 2013/mai	37%
¹⁰ 2012) 130 Also, a second table illustrates the valence patterns of the LU, i.e. its combinatorial	40.3 1.4 0 3€ 0	36 Criteria fo 2013/mai	37%
11 282 5.3.1.3. Differences in the valence patterns		2 ^c Criteria fo 2013/mai	
12 130 Figure 21. Lexical entry report of argue: valence patterns (FrameNet 2012)		3º Criteria fo 2013/mai	
13 constituents are NP + VP + ADVP. Table 6. Valence patterns of TIDE Table 7. Valence patterns of		5' Process-o 2012/mai	
(who is acting and who is acted upon) and sign (or valence, polarity; is the relation positive or negative?) or		35 Semantic 2011/nov/	
+ ADVP. Table 6. Valence patterns of TIDE Table 7. Valence patterns of MAREA (1a) TIDE to occur	4.19 18 18 0 5' 0	5' Process-o 2012/mai	50%
of the FE is further enriched with the specification of valence patterns, the semantic and syntactic templates	4.11 18 48 0 50 0	5(Process-o 2012/mai	49%
¹⁷ never occur in the contexts of determinar2. Among the valence patterns that the term require1 admits the only	83.9 3. 180 78 0	7t Criteria fo 2013/mai	74%
¹⁸ passive sentences, can also be stated in terms of the valence or dtrs attributes. This is discussed in Section	70.8 2. 9, 0 7(0	7(A CONST 2011/nov/	1359
¹⁹ actantial structures of the verbs revealed that the verbs' valence patterns are different. For instance, commit2,	83.6 3. 9t 0 7t 0	7t Criteria fo 2013/mai	73%
20 . This chapter further enriches the relations with valence or polarity, creating a directed and signed	51.4 1.472 8: 0	6' Semantic 2011/nov/	1199
²¹ meanings specific of this field as well as a syntactic valence or combinatory value. The concentration of such	1.13 54 92 0 21 0	2' SEMANTI 2012/mai	21%
22 meanings specific of this field as well as a syntactic valence or combinatory value. Naturally, such noun	476 2! 94 0 1E 0	1€ Frame-Ba 2012/mai	15%
²³ lexical and a constructional description of the TAM and valence patterns resolve the apparent inconsistencies of	29.3 1.7(0 3) 0	3! Constructi 2011/nov/	73%
24 realizations of the different FEs, as well as their valence patterns or mappings between semantic and	3.50 1677 0 4: 0	4: Process-c 2012/mai	42%
²⁵ , adverbs, and prepositions as well as to identify their valence patterns (the ways in which the semantic	39.0 1.74 0 38 0	3{ Criteria fo 2013/mai	36%
26 of determinar2. The Portuguese verb admits three valence patterns: JUDGE (Sub. NP) PROTAGONIST	83.8 3.7(0 7! 0	7t Criteria fo 2013/mai	73%
, resolve1 and suprir1. 5.3.1.3. Differences in the valence patterns About 6% of the total number of	83.5 3. 1(0 7! 0	7! Criteria fo 2013/mai	73%
²⁸ and issues? RQ1c Can we automatically determine the valence (positive, negative) of these semantic relations?	4.74 18 57 0 2' 0	6 Semantic 2011/nov/	9%
²⁹ construction to translate this phrase although there is no valence reduction apparent in the Abenaki. Leavitt		3' The Morpl 2011/nov/	63%
30 : he'elim In languages that morphologically mark valence reduction, these unaccusatives often bear	2.91 1060 20 0	2f Hidden en 2013/mar	26%
31 discusses the applicative suffix ,d□. 3.6 Summary of Valence Reducing Suffixes The most important		3€ VALENCE 2011/nov/	74%
³⁷ verbs do not appear in a morphological form typical of valence reducing operations: (14) ha-ra'ayon xamak	3.02 10 30 21 0	27 Hidden en 2013/mar	27%
33 for different reasons and with different results. Table 3. Valence Reducing Suffixes Suffix Description Function	10.4 7:2:0 3:0	37 VALENCE 2011/nov/	74%
in section 3.2.1.3. For this reason, the sense and valence rich descriptions provided by FrameNet and		36 Criteria fo 2013/mai	
25 Crime Safety Public Order & Safety Crime Justice Valence Rule of Law Economic Growth Economic		68 Semantic 2011/nov/	
njn ?kd qm?. Each suffix indicates an increase in valence, resulting in a net addition of two 104 arguments		84 VALENCE 2011/nov/	
³⁷ participant a ectedness marking, to a set of voice and valence-related alternations. However, all of these	64.5 2.960 86 0	86 Constructi 2011/nov/	: 1699
concordance collocates plot patterns clusters timeline filenames source text notes		F/ 1 00007 00///	
347 entries Row 7 bject) (Agreement= Verb root -Valence -Scenario -Reflexive/			

Figure 14. Concordance lines for the cluster verb root valence.

Now you can analyze this cluster according to your research.



Although WST is a paid program, the author (Mike Scott) makes the version 5 available for free in his site: https://lexically.net/wordsmith/

⁷ If we would use the n-gram concept, this would be a trigram.

MÓDULO 2

Corpus Linguistics: Teaching and Learning

Basic Contents

- At school, at home, self-learning
- Online corpora

Objectives

- •Realize that Corpus Linguistics is a useful tool for teachers and learners.
- Access a lot of free tools available on the Internet.

CORPUS LINGUISTICS: TEACHING AND LEARNING

ACTIVITY 5

Video class, module 2. Watch the professor's hints about the subjects that are going to be worked in this module.

At school, at home, self-learning

Corpus Linguistics can be a very powerful teaching and learning tool. As a teacher, you can show your students how to solve their linguistic problems by themselves. As a learner, you can learn a lot of language patterns and also solve your doubts. Differently from other areas in Linguistics, Corpus Linguistics tools and texts are generally available at free cost on the Internet.

The book **Corpora no ensino de línguas estrangeiras** brings many ideas how corpora can be worked to teach. Read a review of this book in here: http://www.seer.ufu.br/index.php/dominiosdelinguagem/article/view/12335

We're going to present, in this module, some of the tools you can freely use on the Internet and how to interact with them.

COCA

The Corpus of Contemporary American English is the biggest freely available corpus on the Internet. Nowadays, with 450 million words (it gets more 25 million words every year), it presents a very easy and intuitive interface. From there, you can also access other corpora using the same search structure.

Something very important about COCA is the balance among its texts.



Figure 15. Registration corner. Source: COCA.

At the first page of COCA, there are links for many other sub corpora, specific features of the system and downloads you can try:

🙆 🛃 Corpus of Con	temporary American English 🕞 🚯 🛈 📄 🕖 🧾 📙 🖽 🕐 🕐
	Figure 16. COCA's upper menu.
	Table 2. Features related to COCA.
Word and Phrase (analyze texts)	Enter entire texts and see detailed frequency information on the words in the text and create word lists based on your text. Click through the words to see detailed information on any word. Highlight phrases in your text and have it search for related phrases in COCA. Search and browse the most complete frequency dictionary of English. See detailed information (all on one page) definition, frequency by genre, collocates (nearby words), concordance lines, synonyms, and Wordnet-related words, all with useful links from one resource to another.
Word Frequency	You can also download lists showing the frequency of the top 60,000 lemmas by genre (and sub-genre), as well as the top 200-300 collocates (nearby words) for these lemmas (4,800,000 node/collocate pairs). There is also a free list of the top 5,000 lemmas in COCA. And now you can download the 100,000 integrated word list from COCA, COHA, BNC, and SOAP the largest, corrected frequency list of English.
Collocates	Download lists with the top 200-300 collocates (nearby words) for 60,000 different lemmas 4,300,000 node/collocate pairs in all.
<u>N-grams</u>	Download free lists containing the top 1,000,000 2-grams (two word sequences), 3-grams, 4-grams, and 5-grams in COCA. There are also other lists that contain the frequency of all 2, 3, and 4-grams (up to 155 million rows of data).
Academic vocabulary	Download free lists containing "core" academic words in 120 million words of COCA-Academic texts (including grouping by word families), as well as the top 20,000 words overall in COCA-Academic. See <i>Applied Linguistics</i> article, or compare to the Academic Word List (Coxhead, 2000).
<u>Word and Phrase</u> (academic)	Similar to the two Word and Phrase resources below, but limited strictly to the 120 million words of academic texts in COCA. Get detailed information on words and phrases, frequency by sub-genre (e.g. Law, Medicine, Science, Business, Humanities), and concordances and collocates in just the academic text. Also, analyze entire academic texts.

On the left upper left site, you find the search display to star your queries (Figure 17).

List Chart Word Browse Collocates Compare KWI	с -
[POS] ?	
Find matching strings Reset	
E Sactions, Toute Mittual, Sart/Limit, Options	
Sections Texts/Virtual Sort/Limit Options	
Figure 17. COCA's search display.	

Using this search display, let's try a basic query? You're producing a text and suddenly a doubt comes out: we use at the Internet, in the Internet or on the Internet? You know this is a kind of problem the grammar books are not going to solve. Instead of looking for a native speaker, you can search the three possibilities using the COCA. Click on LIST, write down the possibility you want to analyze (for example, at the Internet) and click the FIND MATCHING STRINGS button. What you get (figure 18) is a screen on the right side, with the frequency found (in this case, 155 entries); note that the middle menu changed from the first to the second option (SEARCH > FREQUENCY).

! 🙆 (C	Corpus of Co	ntempor	rary An	nerican Er	nglish [i 🖹 🕓	
	SEA	RCH	I	FREQUEN	CY		CONTEXT		ACCOUNT
ON CLICK:	CONTE	XT 🚱 TRANSLATE (??)) G GOOGLE		PRON/VIDEO	BOOK (HE	ELP)		
1		AT THE INTERNET					155		
									0.422 sec

Figure 18. COCA's query result.

If you click on the combination you are searching (at the Internet), you get a KWIC screen with all the concordances (figure 19); again the middle menu changed, now from he second to the third option (FREQUENCY > CONTEXT):

!	G		Gorpus of Co	n	ten	nporary American Ei	nglish 📴 🚯 🛈 🖹 🤇	▶ 📕 💶 🖽 🕐 🥐
			SEARCH			FREQUENCY	CONTEXT	ACCOUNT
		_E: <u>100</u> < 1/2	> >>					
CLIC	K FOR	MORE (CONTEXT]	SAV	E LIST CHOOSE LIST CREATE N	EW LIST [?]	SHOW DUPLICATES
1	2019	FIC	Obsidian	А	ВC	memory. It was one of the many words	ne wanted to research <mark>at the Internet</mark> caf when he re	turned home. # The Black girl talked about the same
2	2018	SPOK	NPR_Sunday	A	BC	thing where they're either like super pre-	cocious because they're so good <mark>at the Internet</mark> and	they're like YouTube billionaires or they're hopeless a
3	2018	MAG	Ars Technica	A	ВC	Bioglio et al. applied their algorithm to i	nformation collected on 47,000 films <mark>at the Internet</mark> M	Movie Database (IMDB). (The authors caution that the
4	2017	SPOK	NPR_FreshAir	A	BC	a lot of people don't like him. And when	I looked <mark>at the</mark> internet, I discovered there is this bur	nch of people that don't like you.
5	2017	SPOK	CNN: Primetime Justice	А	вс	type of privacy issues. You have so ma	ny young people look <mark>at the Internet</mark> as, This is my p	rivate thing, and they don't think about everyone
6	2017	MAG	A.V. Club	А	вс	this sense it is perhaps the perfect thing	to listen to while looking at the internet today. Don't	say we didn't warn you.
7	2017	MOV	After Porn Ends 2	А	вc	and then it destroyed the business pret	y much. People can just look <mark>at the Internet</mark> and get	whatever they want for free. There's no need to put $\boldsymbol{\epsilon}$
8	2016	MAG	TechCrunch	А	ВC	New rules in the era of " things " # Let's	look at the Internet of Things. This is the hardware in	dustry of the future, and it will
9	2016	MAG	Fortune	А	ВC	competitors, Verizon Communications v	z and AT&T t, are also taking aim <mark>at the Internet</mark> of th	ings opportunity. The bigger players are aiming to es
10	2016	NEWS	Washington Post	А	ВC	Drudge was feeling celebratory on Twitt	er. # (Drudge, no dummy <mark>at the Internet</mark> , purges his	tweets regularly, which is why we've included a screei
11	2015	SPOK	ABC: The View	А	ВC	internet companies. Is there a responsib	ility for someone to take a look <mark>at the internet</mark> and lo	ook at it for what it is, which is a tool for spreading
12	2015	MAG	NatGeog	А	вc	York City. Adeoti makes twice the salary	he made as the manager <mark>at the Internet</mark> cafe. But all	this exposure to money and movies had whetted his
13	2015	ACAD	QuartRevDistanceEd	А	ВC	: What do students think, want, and do?	Paper presented at the Internet Research 14.0, Denv	er, CO. # Dennen, V. P., & Burner,
14	2015	ACAD	QuartRevDistanceEd	А	вc	and the higher education classroom: Stu	ident preferences and attitudes. Paper presented at	he Internet Research 15.0, Daegu, South Korea. # Di
15	2015	τv	The Amazing Race	А	вС	are the Olympians We were going dov	nstairs To look <mark>at the Internet</mark> And we randomly	ran across The Olympians. We're going
16	2014	SPOK	PBS_Newshour	А	вc	're really more volatile because of the ex	pectation, particularly when you look at the Internet	stocks, then sort of new technology, social media stc
17	2014	FIC	Ploughshares	А	вс	chances of remission were excellent, ev	eryone said. " Don't look <mark>at the Internet</mark> , " her new or	cologist warned her, and Merel, numb and childlike v
18	2014	FIC	WarLitArts	А	вс	been contracted out from elsewhere. Th	e guards at the chow hall and <mark>at the</mark> internet cafe we	ere Ugandan Army, many of whom had seen action d
19	2014	MAG	PopMech	А	вс	Past eras simply can't compete. # For pr	oof, just look <mark>at the Internet</mark> . Nobody could have ima	gined typing something and everyone in the world b
20	2013	SPOK	CNN: CNN Live Event	A	ВC	. So certainly, because there are around	two million citizens who look at the internet , the gov	ernment is allowing this discussion to happen. So, it

Figure 19. COCA's concordance lines for the query at the Internet

As you can see, the platform gives you a lot of information about each line of concordance. In line 1, for example, we have: the year the material was published (2019), the media it was published (fiction), the name of the media (Obsidian). If you click on this links, you get part of the text with the information highlighted (figure 20); once more the middle menu changed, this time from the third to the fourth option (CONTEXT > CONTEXT+):

! 🙆 🚱 Corpus	s of Contemporary American English		
SEARCH	FREQUENCY	CONTEXT	CONTEXT +
Source information:			
Source	FIC: Obsidian: Literature in the African Diaspora		
Date	2019 (2017/03/22)		
Publication information	Vol. 43, Issue 1		
Title	PLASTIC CITY		
Author	Leslie Ann Murray		
spot that most needed massaging. # until it lodged into his memory. It was about romantic relationships, famil	paned and then worked on her lower back. He moved his hands arounn " Men suck, " she said. " I wish I was not committed to heterosexuality. s one of the many words he wanted to research <u>at the Internet</u> caf whe y life, work life and life. Her present stress was Marius-related. Whenev y oyou think? " The Black girl asked. " Are you listening to me? " # " Um, k	" # " Heterosexuality. " He did not have n he returned home. # The Black girl t er tourists started talking about this pe	e his notebook, so he quietly repeated the word alked about the same things all tourists talked ersonal side of their lives, he blocked them out
Source information:			
Date Publication information Title Author Source	2017 Jan/Feb 2012 Net Workh Negulescu, Kns Carpenter Lechnology keview		
among like minded communities can we hope to preser NEGULESCU IS DIRECTOR OF WEB ARCHIVING AT THE	But such cooperation is appearing within international research communities and cultural groups in both i ve any significant size of the Web. The future dees not afford anyone the luxury of the unlimited time, INTERNET ARCHIVE, A NORPORTI INTERNET LISBARY THAT PRESERVES DIGITAL CONTENT 920030601 # Proceedings of the 24th ACM Symposium on User Interface Software and Technology, Santa Bathare,	funds, computing power, and storage capacity that woul NFORMATION TECHNOLOGY # Touch screens that work to	ld be needed to do it alone. AuthorAffiliation # KRIS CARPENTER through fabric # SOURCE: * POCKETTOUCH: THROUGH-FABRIC

Figure 20. Expanded concordance line for the first example.

ACTIVITY 6 - SEARCH

Using the COCA, search the other possibilities given above using a *preposition* + *the Internet*. What are your findings? Can you systematize a rule of usage?

http://

ACTIVITY 7 - DISCUSSION

Discuss with the other students, using the AVA forum, what is the best way to use and the differences among the structure: *preposition* + *the Internet* given above. The participation

in the discussion will be evaluated.

The activities we performed before are just one among a lot of others we can try using the COCA platform. Take a look at possible searches using combinations and wildcards in table 3:

Type of search	COCA-General
Specific word or phrase	l guess
Substring	*al_j
Lemma (forms of a word)	CONJ PRON BE like , (<i>and she was like ,</i>)
Part of speech	ADJ eyes
Synonyms	=strong
User-defined lists	@colors @clothes
Sortable concordance lines	fathom
Collocates (nearby words)	BREAK_v
Use Mutual Information score	BREAK_v
Compare two words	utter / sheer

Table 3. Query possibilities using the COCA interface.

Let's try, now, another important feature of COCA: the Word and Phrase tool (figure 21). Access using the icon.

		You can always get to this page by clicking this icon above.
		You can enter any text that you would like in the form at the left – for example a paper that you've written, or a newspaper article that you've copied from another website. After inputting the text, you can then see useful information about words and phrases in that text, based on data from COCA.
SEARCH		First, it will highlight all of the medium and lower-frequency words in your tex and create lists of these words that you can use offline. This frequency data car help language learners focus on new words, and it can allow you to see "what the text is about" (i.e. text-specific words).
HELP	Compare to previous WordAndPhrase (PDF)	Second, you can click through the words in the text to see a detailed "word sketch" of any of the words showing their definition and their translation (ir
	Overview	more than 100 languages); links to pronunciation, images, and videos; related topics, collocates, "clusters" (2, 3, and 4-word phrases); and concordance lines.
Þ	Word-oriented functions	Finally, you can do powerful searches on selected phrases in your text, to show
	Phrase-oriented functions	related phrases in COCA. In this way, this resource is like a "collocationa thesaurus" to see what related phrases are most likely in different styles o English.
		Just enter some text, and there will be more help files on the next page.

Figure 21. Word and Phrase main page. COCA platform.

Let's try a simple word first. LINGUISTICS is the word we are going to search. Just write the word and click the SEARCH button.

Corpus Linguistics: Teaching and Learning

DIT TEXT	SAVE TEXT	• WORD	PHRASE	(CLICK ANY WORD FOR	FULL WORD SKETCH)	
				LOW FREQ	MID FREQ	HIGH FREQ
FREQ RANGE			> 3000			
1 WORDS			100 %			
	D BELOW FOR A FUI			1: linguistics		

Figure 22. Frequency query for Linguistics.

What do you get in this screen? On the left side, our search word (Linguistics) is classified, according to its frequency, among three levels of words in the COCA corpus: range 1-500 (in blue, extremely common words), range 501-3000 (in green, common words) and range above 3000 (in yellow, rare words or words belonging to specific domains). As you can see, LINGUISTICS belongs to the third range, i.e.⁸, it's not a very common word in English (it doesn't have a very high frequency in the corpus). Now, let's learn more about it: click on the word linguistics (in yellow, left corner). What we get is a complete description of this word in the English language (figure 23). Click on the innumerous links to learn more about it.

linguis	tics (NOUN) 🗙 🔨 #15174	TOPICS (more)
BLOG WE		inguistic, language, eg, grammar, ie, linguist, english, usage, vocabulary, discourse, correspond, empirical, similarity, speaker, verb, grammatical, literature, methodology, namely, english
language and	fic study of language 2, the humanistic study of l literature DMOCG E JGlish PlayPhrase Yarn	NOUN professor, university, language, study, association, corpus, department, field VERB study, teach, combine, reconstruct, specialize, invent, portray, practice
SYNONYMS (I	more)	ADj applied, computational, cognitive, historical, clinical, modern, literary, arabic ADV eg, primarily, surely, genetically, per, se, the, flat
semantics, sy		
	ore)	guistics department + linguistics from + linguistics to + linguistics with + linguistics has
semantics, sy	ore) linguistics at + linguistics professor + linguistics in + ling	guistics department + linguistics from + linguistics to + linguistics with + linguistics has uistics + corpus linguistics + historical linguistics + cognitive linguistics + clinical linguistics + to
semantics, sy CLUSTERS (mi inguistics •	ntax ore) linguistics at * linguistics professor * linguistics in * ling in linguistics * applied linguistics * computational ling linguistics	
inguistics • inguistics •	ntax ore) linguistics at + linguistics professor + linguistics in + linguistics = applied linguistics = computational linguistics linguistics professor at + linguistics and education + linguistics and literary + linguistics and phonetics	uistics * corpus linguistics * historical linguistics * cognitive linguistics * clinical linguistics * to
semantics, sy cLUSTERS (me inguistics • inguistics • inguistics • • • Inguistics • •	Inguistics at a linguistics professor a linguistics in a linguistics applied linguistics are computational linguistics and education a linguistics and literary a linguistics and phonetics professor of linguistics are for computational linguistics of linguistics a department of linguistics	aistics * corpus linguistics * historical linguistics * cognitive linguistics * clinical linguistics * to aguistics and literature * linguistics and applied * linguistics and literacies * linguistics and philosophy * language and linguistics * degree in linguistics * field of linguistics * in applied linguistics * institute nguage * linguistics and the novel * linguistics and the school * linguistics and evolutionary biology *

⁸ Learn about Latin abbreviations in English: https://en.wikipedia.org/wiki/List_of_Latin_abbreviations

		ublic Policy • ACAD:Style • ACAD:BMC Bioinformatics • WEB:m		
		lied Linguistics and TEFL • ACAD:AI Magazine • BLOG:agelog		
		ratory Phonology: Journal of the Association for Laboratory P		
AD:Style	• ACAD:Style • ACAD:Sty	/le • WEB:cenotes.blogspot.com • BLOG:blogs.suntimes.com	• WEB:tab	iletmag.com • BLOG:reason.com • ACAD:J Am Folklore •
ONCORD	ANCE LINES (more)			
ACAD: 2	017: Al Magazine	Equations . Transactions of the Association for Computational	Linguistics	3 : 585-597. # Koncel-Kedziorski , R. ; Roy , S. ;
2 ACAD: 1	995: AcademicQs	than on religion , economics , sociology , or language and	linguistics.	5 # Will someone inform the professorate that , at a time
B ACAD: 1	996: CurrentPsych	subjects their message would only be used in a computerized	linguistics	analysis . Second , in the audience condition subjects gave the
4 ACAD: 1	994: AmerStudies	; and the non-membership ERIC Clearinghouse on Language and	Linguistics	and Center for Applied Linguistics serve as a pool of research
5 ACAD: 1	992: AmerStudies	American students.72 # Staff and students of	linguistics	and education 🛛 as well as other disciplines , have not been
5 ACAD: 2	017: Al Magazine	first attempts to put together researchers from computational	linguistics	and from argumentation theory.
7 ACAD: 1	998: SocialStudies	archaeology is one of four subfields , which also include	linguistics	and physical and cultural anthropology . Despite its
3 SPOK: 2	009: NPR_TalkNation	tactics here . DAVID : Well , we re also talking	linguistics	and pro-life is - pro-life is anti-choice . Pro-life is - back
9 MAG: 19	990: AmSpect	here . # We are not told either whether he studied	linguistics	at a university or whether it is just his avocation . It
0 NEWS: 2	012: NYTimes	Think Jen Johnson 's keypad thumbs . A graduate student in	linguistics	at Georgetown University , Ms. Johnson , 21 , stumbled onto
1 NEWS: 2	012: NYTimes	, Ms. Johnson , 21 , stumbled onto Siletz while studying	linguistics	at Swarthmore College , which has helped the tribe build its
2 NEWS: 2	012: NYTimes	, " said K. David Harrison , an associate professor of	linguistics	at Swarthmore who worked with the Siletz tribe and the other
3 WEB: 20	12: amazon.com	theories of morality should explain moral behavior , much as	linguistics	attempts to explain human verbal communication . Philosophers
4 BLOG: 2	012: vickiarcher.com	decades of formally studying and/or teaching literature &;	linguistics	before 📔 realized that my East Tennessee grandmother was
5 BLOG: 2	012: reason.com	but as a linguist I need to point out that all	linguistics	can tell us is that the oldest languages we know anything at
6 ACAD: 2	005: AnthropolQ	social distance , rather than occidental academic historical	linguistics	categories 📙 for example , the Greenberg African language
7 WEB: 20	12:tword.dictionary	phoneme (I think it 's been a while since	linguistics	class) in the English language is " er " . #
8 FIC: 201	4: VirginiaQRev	n't have to work for a living . Her degree in	linguistics	combined with her halting English did n't leave her many options
9 BLOG: 2	012: a-sense-of-place.com	lab , information and decision systems lab , and its	linguistics	department . It opened in 2004 after a lot of funding from
0 NEWS: 1	990: WashPost	Worcester , Mass . He received a master 's degree in	linguistics	from Georgetown University and a doctorate in anthropology
1 ACAD: 2	001: AnthropolQ	also how it is interpreted . Recent work in anthropology and	linguistics	has greatly improved our analysis of interpretation by showing
2 WEB: 20	12: deirdremccloskey.com	the cheap talk paradox '') . A tamous story in	linguistics	illustrates the point . A very pompous linguist was giving a talk
3 ACAD: 1	996: AmerStudies	its function ; # c) insufficient development of contrastive	linguistics	in the aspect verging on culture studies ; # d) the
4 WEB: 20	12: amazon.com	perspective of Noam Chomsky 's theories that transformed	linguistics	in the 1960s . Stephen Pinker 's " The Language Instinct "
5 ACAD: 1	995: ArtBulletin	the disciplinary and professional structure of knowledge ?	Linguistics	is a discipline ; English is a department ; cinema studies is
6 NEWS: 2	013: NYTimes	Hamdallah was educated in the West his Ph.D . in	linguistics	is from the University of Lancaster and is not officially tied
	i	Figure 23. Word LINGUIS	TICS	

As you can see in figure 23, the concordance lines present you words with different colors > they represent different word classes. We know that Linguistics is a noun. Can you guess the association among the other colors and word classes?

There are innumerous ways to analyze combinations with Linguistics. Click on the other possibilities (as in figure 24) to get new screens:

```
See in iWeb 🝙 Collocates Clusters Topics Dictionary Texts KWIC 🞍 HELP
Figure 24. Menu to get to new screen
```

Try also to search with a text, like in figure 25. The result (figure 26) can be analyzed word by word or as a whole.

-SAMPLES-

MY TEXTS

America is rich in small businesses. These enterprises account for over 30 million U.S. businesses and some two-thirds of net new jobs. While venture-backed startups generally skew white, male and coastal, these Main Street companies actually look like—and drive—America. To shine a light on these entrepreneurial heroes, Forbes has created the Next 1000. This year-round initiative showcases the ambitious sole proprietors, self-funded shops and pre-revenue startups in every region of the country—all with under \$10 million in revenue or funding and infinite drive and hustle. Fueled by your nominations and screened by top business minds and entrepreneurial superstars, these new faces will number 1,000 by

SEARCH CLEAR

Figure 25. Text search

Corpus Linguistics: Teaching and Learning





Something very important about COCA: they also have Portuguese corpora, using the same interface of the English corpora. Take a look here (figure 27):

https://www.corpusdoportugues.org/xp.asp

FREE Free Free Free Free Free Free Free Fr	o corpus do português								
Os corpora Tamanho dos corpus Comparar Recursos Problemas (inglês)		do pelo Professor Mark Davies, e da coleção corpora da BYU.	BYU. Financiado pelo National Endov	ment for the Hu	English Português manities (2004, 2015). Faz				
Contatar		Corpus	Tamanho	Criado	Mais informação				
	1	Género / Histórico	45 milhões de palavras	2004-06	Info				
Humanities	2	Web / Dialetos	1 <i>mil milhão</i> de palavas	<u>2015-16</u>	Info				
BYU	3	NOW (2012 - 2019)	1,1 <i>mil milhão</i> de palavras	2018	Info				
	4	WordAndPhrase	40.000 palavras principais	2017	Info				
	•		, que permite ver as mudanças histório ior) que permite verificar as variaçõe:						

Figure 27. Corpus do Português, COCA platform.



ACTIVITY 8 - PIPE

It's your time to evaluate a site that works with corpora. What's the idea? You find a site that has tools to work with corpora already compiled (like Projeto COMET, BNC, other COCA platforms, Linguateca, LAEL, Webcorp, etc.) and you're going to describe the site, as we have done with COCA above. You can also print the screens if you want. This work will be evaluated by your tutor.

Lexical Analysis Tools

Basic Contents

- Lexical analysis software overview.
- Corpus compilation.
- AntConc suite for corpus analysis.

Objectives

- Have the basic knowledge about how to compile a corpus.
- •Use the basic tools (wordlist, keywords and concordance) of a lexical analysis software.

LEXICAL ANALYSIS TOOLS

ACTIVITY 9

Video class, module 3. Watch the professor's hints about the subjects that are going to be worked in this module.

There are many software you can use to describe language. The most modern and complete ones must be bought (like WordSmith Tools) or subscribed (like the Sketch Engine).

Although old, the following text shows some differences among them: http://www.ileel. ufu.br/guifromm/upload/ferramentasdeanaliselexicalcomputadorizadas.pdf. For a more updated text, comparing WordSmith Tools and Sketch Engine, try this one: http://www. periodicos.letras.ufmg.br/index.php/relin/article/view/15766

An example of a very powerful corpus analysis suite is the WordSmith Tools (http://lexically.net/wordsmith/index.html):

		R C R R R R R R R R R R R R R R R R R R				
Previous results	On startup	Update check				
🎯 Main settings	remember screen position					
🚐 Print settings		weekly · check now				
Colour settings	restore last work saved	Version 8.0.0.48 (04/02/2021)				
🧮 Folder settings		latest version =				
🔁 Language settings	show Help					
Concord						
📉 KeyWords	show toolbar in Tools					
祝 WordList	[∞] show statusbar in Tools					
💦 WSConcgram	show statusbar in roois					
💽 Chargrams	type/token ratio is too sensitive to corpus size					
dtilities						
About						
		A Advanced Settings				
	System					
	Windows default ·	Associate/clear file extensions				
	Get Started Guide	Support				
Portuguese	0% 0%	< > Help				

Figure 28. WordSmith Tool's main menu.

There are the three main tools, very common in other suites: wordlist, keywords and concordance. But WordSmith Tools has a lot of other smaller programs:

	KeyWords	WordList
		e ve all settings
revious results	Utili	ties
Main settings		
Print settings	Aligner	Character Profiler 🛛 😡
Colour settings		
Folder settings	CharGrams 🛛 🔒	Corpus Checker 🛛 💽
Language settings		
Concord	File Utilities	File Viewer 🤤
KeyWords		
WordList	Language Chooser 🛛 🚱	Minimal Pairs 🛛 😽
WSConcgram		
o Chargrams	Registration	Text Converter 🛛 🔂
e Utilities		
bout	WordSmith version check 🛛 🚾	WSConcGram

Figure 29. WordSmith Tools utilities.

In our course, we're going to work with a free option. The most famous one is the AntConc suite (http://www.laurenceanthony.net/software.html):

Lexical Analysis Tools

🍓 AntConc 3.5.9 (Window	rs) 2020						_		\times
File Global Settings Too	Preferences H	lelp							
Corpus Files	Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates W	ord List Keyv	vord List		
	Concordance H								
	Hit KWI	3					File		^
		Words 🗌 Case				Search Win			~ ~
	Search Term			Advanced		50			
Total No.	<u> </u>	~	c .						
0	Start Kwic Sort	Stop	Sort	Show Every Nth F	Row 1				
Files Processed	Level 1 1R	Level 2 2R		Level 3 3R 🔹			(Clone Res	sults

Figure 30. AntConc's main menu.

As AntConc is free, we're going to use this program to demonstrate the capabilities of describing language. But first, we must compile a corpus.

Compiling a corpus

Following the idea presented on module 1, let's create our corpus. The summary of it:

Language	English
Sources	Academic Texts (articles, MA's and PhD's thesis)
Time	Synchronic
Selection	Study Corpus
Content	Specialized – Linguistics, Prosody
Authorship	Native and non-native speakers
Internal Distribution	monolingual
Size	Small
Codification level	Header

Table 4. Corpus compilation.

We propose you to create a corpus in a subfield of Linguistics, Prosody. Let's see the steps:

Finding the texts

The best place to find texts is Google. But pay attention, a simple query wouldn't be enough. Use this kind of search:

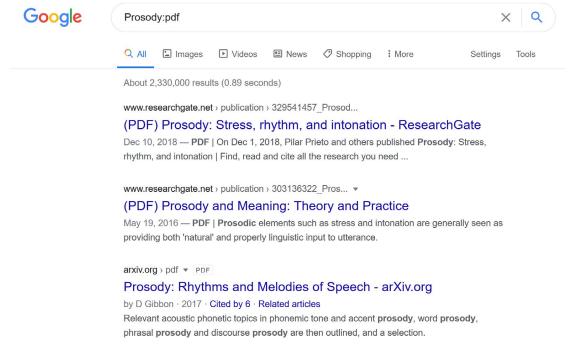


Figure 31. Google: PDF query for the word Prosody.

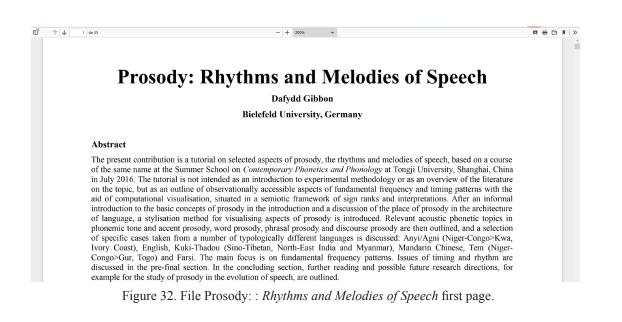
Take a look at the upper right corner of picture 22. Let's click the *options* menu and choose *advanced search*. We get the following screen:

We choose the pdf extension because the majority of articles (from specialized journals) and thesis available on the Internet are in this format.

Organization is very important when we work with Corpus Linguistics. To easily find the files you're going to work with, create a directory, in your computer, named Corpus Linguistics.

Let's take the third example. When you click the link, the site opens a file (named Prosody: Rhythms and Melodies of Speech). Before saving it, open your Corpus Linguistics directory and create a subdirectory named Corpus PDF. Save the file in it⁹. Notice that the file has 35 pages.

⁹ Instead of using the file original name, try a shorter classification, like: File 1, File 2, etc.



In the sequence, save other files into your directory.

Something very important about the files you're going to use with lexical analysis tools is the format: the best way to work with the programs is to save your files in .txt format. Let's do the following: 1. Create a new subdirectory in your Corpus Linguistics directory, named Corpus TXT; 2. Open the Windows menu and choose All the programs > Windows Accessories > Notepad (figure 33); 3. open the Prosody: Rhythms and Melodies of Speech file, click the keys control + A together (to select the whole text, figure 34), copy it and paste it in the notepad file (figure 35).



Figure 33. Notepad

	is 1704.02565.pdf ×		
🗘 ዋ 🖥	} ⊜	(h) (i) 1 / 35 (h) (ii) (iii) (i	
	-		
	Proso	ody: Rhythms and Melodies of Speech	1
		Dafydd Gibbon	
		Bielefeld University, Germany	
	Abstract		
4	of the same name at the in July 2016. The tutor on the topic, but as an aid of computational introduction to the bas	tion is a tutorial on selected aspects of prosody, the rhythms and melodies of speech, based on the Summer School on <i>Contemporary Phonetics and Phonology</i> at Tongji University, Shanghe torial is not intended as an introduction to experimental methodology or as an overview of the an outline of observationally accessible aspects of fundamental frequency and timing patterns al visualisation, situated in a semiotic framework of sign ranks and interpretations. After an asic concepts of prosody in the introduction and a discussion of the place of prosody in the arcl sation method for visualising aspects of prosody is introduced. Relevant acoustic phonetic t	ai, China literature with the informal hitecture

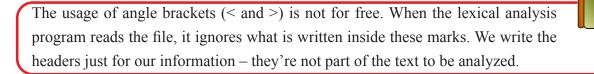
Figure 34. File Prosody: Rhythms and Melodies of Speech - whole text selected and ready to be copied.

As a result, you get this screen:



Figure 35. file Prosody: Rhythms and Melodies of Speech copied to the Notepad¹⁰.

Now we're going to insert a header (figure 36) for the file, containing its internet address and the date we collected the text. You're going to do the same procedure with all



¹⁰ As you can see, the Notepad erases all the formats, formulas, figures, etc. from the original text. Don't worry, because they're not important for corpus analysis.

Finally, you get the file complete:



Figure 36. Text from file Prosody: Rhythms and Melodies of Speech copied to Notepad (with header).

The final step is saving the new file. Pay attention to some details: we're going to save the original file (Prosody: *Rhythms and Melodies of Speech*) as File 1:

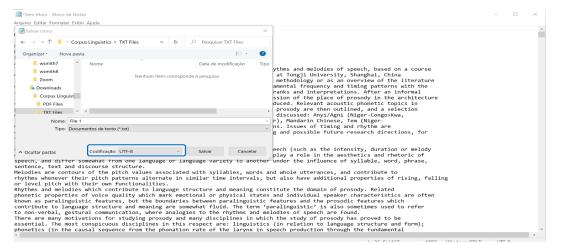


Figure 37. Saving the file Prosody: Rhythms and Melodies of Speech as File 1 and with UTF-8 codification.

Pay attention to the code (codificação) we choose: UTF-8 (figure 37). It's very important to choose this one, because it works better with AntConc. If you want a more concise text, you can erase the pre (abstracts) and post (references) texts¹¹.

¹¹ Generally, they're not important for your research.

AVA

Now you're going to finish compiling the corpus we've started above. The first text is already there (File 1). As soon as you finish compiling it, you are going to compact (into a .zip file) the directory (CORPUS TXT) and send it to your tutor. Your corpus must have, at least, 300,000 words¹²

Don't worry if you don't know how to count the words in different files. The lexical analysis tools will do it. We're going to learn how to use the AntConc in the sequence.

AntConc

You've already seen the main menu of AntConc. As in all lexical analysis software, the first tool we're going to work with is the Word List.

Word List

The first step to work with the Word List is to choose its tab and, in the sequence, open the directory with the files we're going to work with. In this case, let's open the Prosody directory (in your case, don't forget you saved your files in the Corpus TXT subdirectory):

a A	AntConc 3.4.1w (Windows	s) 2014 —	X
File Global Settings Tool Preferences	s Help		
Open File(s) Ctrl+F	cordance Plo File Viev Cluste	ers/N-Gram Collocate Word List K	eyword Lis
Open Dir Ctrl+D	Word Tokens: 0	Search Hits: 0	
Close Selected File(s)	lord	Lemma Word Form(s)	^
Close All Files			
Clear Tool			
Clear All Tools			
Clear All Tools and Files			
Save Output to Text File Ctrl+S			
Import Settings from File			
Export Settings To File			
Restore Default Settings			
Exit	_		
		<	> ~
Search Term	Words Case Regex	Hit Location	
		Advanced Search Only	0
Total No. Start	Stop Sort	Lemma List	Loadec
0 Sort by Inve	ert Order		
Files Processed Sort by Freq	\checkmark	Clo	ne Results

Figure 38. AntConc. Opening a directory in the WordList tool.

¹² You can use a text editor to calculate the size of your corpus. Put all of them in a single file and calculate it (in Microsoft Word, use Revision > Word Counting). File Global Settings Tool Preferences Help **Corpus Files** Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List File 1.txt Word Types: 0 Word Tokens: 0 Search Hits: 0 File 2.txt Rank Freq Word Lemma Word Form(s) File 3.txt File 4.txt File 5.txt File 6.txt File 7.txt File 8.txt File 9.txt File 10.txt File 11.txt File 12.txt File 13.txt File 14.txt File 15.txt File 16.txt File 17.txt File 18.txt File 19.txt File 20.txt File 21.txt File 22.txt File 23.txt Search Term 🗹 Words 🗌 Case 🗌 Regex **Hit Location** File 24.txt Search Only Advanced 0 -Lemma List Loaded Sort Start Stop Total No. Word List Loaded 24 Sort by Invert Order **Files Processed** Sort by Freq **Clone Results** V

When you select to open the directory, you get the names of the files (included in this directory) you chose on the left field. In our case, the directory contains six files:

Figure 39. AntConc. Wordlist and files.

With the files opened, just click the START button (Figure 39). The program processes the corpus and gives you some information:

AntConc 3.5.9 (Window					- 🗆	\times
File Global Settings To	ol Preference	ces Help				
Corpus Files	Concor	dance Cor	cordance Plot File View Clusters/N-	Grams Collocates Word List Keywo	rd List	
File 1.txt		pes: 1731		Search Hits: 0		
File 2.txt	Rank	Freq	Word	Lemma Word Form(s)		^
File 3.txt File 4.txt	1	27095	the			
File 5.txt						
File 6.txt	2	17162				
File 7.txt	3	14652	a			
File 8.txt	4	11127	and			
File 9.txt	5	10759	to			
File 10.txt	6					
File 11.txt File 12.txt		10676				
File 13.txt	7	9288	is			
File 14.txt	8	7428	that			
File 15.txt	9	5694	e			
File 16.txt	10	5122	s			
File 17.txt						
File 18.txt File 19.txt	11	4773	t			
File 20.txt	12	4329	i			
File 21.txt	13	4297	as			
File 22.txt	14	3841	for			
File 23.txt	15	3560	be			
File 24.txt						
	16	3324	it			
	17	3321	are			
	18	3112	this			
	19	2992	n			
	20	2954	not			
	21	2778	or			
	22	2733	with			
	23	2580	0			
	24	2579	by			
	25	2449	r			
	26	2400	on			
	27	2340	an			
	28	2120	which			
	< >	< >	<	> <		> 🗸
	Search	Ferm 🗹 W	Vords 🗌 Case 🗌 Regex	Hit Location		
			Advanc	ed Search Only 0		
	C		Chan Cast	Lemma List Loaded		
Total No.	Star	L.	Stop Sort	Word List Loaded		
24 Files Processed	Sort by	Invert	Order			
riles riocessed	Sort by	Freq	~		Clone	Results

Figure 40. AntConc. Corpus of Prosody, processed by the Word List tool.

In the middle of the screen, you find a lot of information: **Word types**, **Word Tokens**, and the **words** the corpus contains. What does all this information mean? It means we have a total corpus of 474,727 words (tokens), with 17,316 different words (types) and the most common word of this corpus is the article THE (it appears 27,095 times in the corpus).

These are the procedures you're going to follow when you are compiling your corpus and you want to know how many words you've already compiled.

Let's work a little bit more with the list. Suppose we're just looking for content words (nouns, adjectives and verbs). There's a way to reduce this list by using a Stoplist. A Stoplist contains words that you don't want the program to analyze. Usually we must prepare a list of words to achieve this goal, but here, we have already created one (download here: https://drive.google.com/file/d/0B_pmpE08GOg2VIZVS0wzeTIyUzA/

Don't forget to save this file in the Corpus Linguistics directory of your computer. Criate a new subdirectory named Lists.

The first step in order to open a Stoplist is to access the Tool Preferences menu > Word List > Use a stoplist below > Open > Add words from file:

۲	Tool Preferences – 🗖 🗙
Category	Word List Preferences
Concordance Clusters/N-Grams Collocates Word List Keyword List	Display Options ☑ Rank ☑ Frequency ☑ Word ☑ Lemma Word Form(s) Other Options ☑ Treat all data as lowercase □ Treat case in sort Lemma List
	Load Load Load Treat Word List Range as Lemma List Range
	O Use all words ○ Use specific words below ● Use a stoplist below
	Add Word Add
	Add Words From File English Stop List - AntCc Open
	a Clear Clear able about above
	according
	Apply Cancel

Figure 41. AntConc. Tool Preferences menu.

Get the file you've just downloaded (English Stop List – AntConc) and click APPLY. Now, run the Word List again (just press the START button). You get a new list:

File 3.txt Freq Word Lemma Word Form(s) File 3.txt 1 9288 is File 4.txt 2 3560 be File 5.txt 2 3560 be File 6.txt 3 3321 are File 6.txt 3 3321 are File 8.txt 4 1656 can File 9.txt 5 1558 have File 10.txt 6 1118 has File 11.txt 6 1118 has File 12.txt 7 1090 language File 13.txt 8 1066 speaker File 14.txt 9 1060 discourse File 15.txt 9 1060 pragmatics File 16.txt 11 1015 context File 18.txt 11 1015 context File 19.txt 12 971 meaning File 20.txt 14 865 speech File 23.txt 15 840 wase	
Rank Freq Word Lemma Word Form(s) File 3.btt 1 9288 is File 3.btt 2 3560 be File 4.btt 3 3321 are File 5.btt 2 3560 be File 5.btt 3 3321 are File 5.btt 1 558 have File 1.btt 6 1118 has File 1.btt 6 1118 has File 1.btt 7 1090 language File 1.btt 7 1090 language File 1.btt 9 1060 discourse File 1.btt 10 1060 pragmatics File 1.btt 11 1015 context File 1.btt 12 971 meaning File 2.btt 14 865 speech File 2.btt 14 865 speech File 2.btt 16 832 pragmatic 17 827 will 18 823 20 805	
File 3.btt 1 9288 is File 4.btt 2 3560 be File 5.bt 2 3560 be File 5.bt 2 3560 be File 5.bt 3 321 are File 5.bt 3 321 are File 5.bt 3 321 are File 5.bt 4 1656 can File 5.bt 5 1558 have File 1.bt 6 1118 has File 1.bt 6 1118 has File 1.bt 7 1090 language File 1.bt 8 1066 speaker File 1.bt 9 1060 discourse File 1.bt 10 1060 pragmatics File 1.bt 11 1015 context File 1.bt 13 870 sentence File 2.bt 15 849 was File 2.bt 15 849 was File 2.bt 15 sea axample	
File 5.btt 2 3560 be File 6.btt 3 321 are File 7.btt 4 1656 can File 8.btt 4 1656 can File 9.btt 5 1558 have File 10.btt 6 1118 has File 12.btt 7 1090 language File 13.btt 7 1090 discourse File 15.btt 9 1060 discourse File 15.btt 9 1060 pagmatics File 15.btt 11 1015 context File 14.btt 11 1015 context File 14.btt 13 870 sentence File 21.btt 13 870 sentence File 22.btt 15 849 was File 22.btt 15 849 was File 23.btt 18 823 example 19 815 use 14 855 20 805 utterance 14 19 815	
File 6.btt 3 3321 are File 7.btt 3 3321 are File 3.btt 4 1656 can File 9.btt 5 1558 have File 1.btt 6 1118 has File 1.btt 6 1118 has File 1.btt 6 1118 has File 1.btt 7 1090 language File 1.btt 7 1090 discourse File 1.btt 10 1060 pragmatics File 1.btt 11 1015 context File 1.btt 11 1015 context File 2.btt 14 865 speech File 2.btt 14 865 speech File 2.btt 15 849 was File 2.btt 15 849 was File 2.btt 15 849 was File 2.btt 15 seample	
File 7.bd 3 3321 are File 7.bd 4 1656 can File 9.bd 5 1558 have File 10.td 5 1558 have File 11.td 6 1118 has File 12.td 7 1090 language File 11.td 6 1118 has File 12.td 7 1090 language File 11.td 8 1066 speaker File 15.td 9 1060 discourse File 15.td 10 1060 pragmatics File 17.td 10 1060 pragmatics File 17.td 11 1015 context File 20.td 12 971 meaning File 21.td 13 870 sentence File 22.td 14 865 speech File 23.td 15 849 was File 23.td 15 829 was If 8 823 example 19 9 815 use	
File 8.bt 4 1656 can File 9.bt 5 1558 have File 10.tt 6 1118 has File 12.bt 7 1090 language File 13.tt 8 1066 speaker File 13.tt 9 1060 discourse File 16.tt 111 1015 context File 16.tt 111 1015 context File 16.tt 111 1015 context File 11.tt 101 1000 sentence File 21.tt 13 870 sentence File 21.tt 14 865 speech File 22.tt 14 865 speech File 22.tt 14 865 speech File 21.tt 13 870 sentence File 22.tt 14 865 speech File 21.tt 14 865 use File 21.tt 823 example 19 815 use use 20 805 utterance	
File 10.txt 5 1558 NaVe File 10.txt 6 1118 has File 12.txt 7 1090 language File 12.txt 7 1090 discourse File 14.txt 8 1066 speaker File 15.txt 9 1060 discourse File 15.txt 10 1060 pragmatics File 15.txt 11 1015 context File 17.txt 11 1015 context File 18.txt 11 1015 context File 21.txt 13 870 sentence File 22.txt 14 865 speech File 22.txt 15 849 was File 22.txt 15 822 pragmatic File 24.txt 15 823 example 19 815 use 11 20 805 utterance 11 21 782 may 11 22 762 information 11 23 731 <td< td=""><td></td></td<>	
File 11.btt 6 1118 has File 12.btt 7 1090 language File 13.btt 8 1066 speaker File 14.btt 9 1060 discourse File 14.btt 9 1060 pragmatics File 15.btt 9 1060 pragmatics File 15.btt 10 1060 pragmatics File 16.btt 111 1015 context File 15.btt 12 971 meaning File 20.btt 12 971 meaning File 22.btt 13 870 sentence File 22.btt 14 865 speech File 22.btt 14 852 pragmatic File 22.btt 15 849 was File 22.btt 16 832 pragmatic File 22.btt 15 849 was File 22.btt 16 823 example 18 823 example set 19 815 use set 20	
File 12.btt 7 1090 language File 13.btt 8 1066 speaker File 14.btt 9 1060 discourse File 16.btt 10 1060 pragmatics File 16.btt 11 1015 context File 18.btt 11 1015 context File 20.btt 12 971 meaning File 21.btt 13 870 sentence File 21.btt 14 865 speech File 23.btt 14 865 speech File 24.btt 15 849 was I6 832 pragmatic I7 827 will 18 823 example 19 815 use 20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do	
File 13.txt 7 1050 language File 13.txt 8 1066 speaker File 15.txt 9 1060 discourse File 15.txt 10 1060 pragmatics File 15.txt 11 1015 context File 18.txt 11 1015 context File 19.txt 12 971 meaning File 20.txt 13 870 sentence File 21.txt 13 870 sentence File 23.txt 14 865 speech File 24.txt 16 832 pragmatic 17 827 will see 18 823 example sex 19 815 use see 20 805 utterance sex 21 782 may sex 22 762 information sex 23 731 like sex 24 724 interpretation sex 25 705 would	
File 15.btt File 15.btt File 16.btt File 16.btt File 16.btt File 16.btt File 18.btt File 20.btt File 20.btt File 20.btt File 20.btt File 23.btt File 24.btt 9 1060 discourse 10 1060 pragmatics 11 1015 context 12 971 meaning 12 971 meaning 13 870 sentence 14 865 speech 15 849 was 16 832 pragmatic 17 827 will 18 823 example 19 815 use 20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
File 16.bt 10 1060 pragmatics File 17.bt 11 1015 context File 18.bt 11 1015 context File 19.bt 12 971 meaning File 20.bt 13 870 sentence File 21.bt 14 865 speech File 23.bt 15 849 was File 24.bt 16 832 pragmatic 17 827 will 18 823 example 19 815 use 20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been Kerth Term V Words □ Case □ Regex Hit Location	
File 17.bt 10 1060 pragmatics File 18.bt 11 1015 context File 18.bt 12 971 meaning File 20.bt 13 870 sentence File 21.bt 14 865 speech File 23.bt 15 849 was File 24.bt 16 832 pragmatic 17 827 will will 18 823 example set 19 815 use set 20 805 utterance set 21 782 may set 22 762 information set 23 731 like set 24 724 interpretation set 25 705 would set 26 676 do set 28 642 been set	
File 18.bt 11 1015 context File 19.bt 12 971 meaning File 20.bt 13 870 sentence File 21.bt 14 865 speech File 23.bt 15 849 was File 24.bt 16 832 pragmatic 17 827 will 18 823 example 19 815 use 20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
File 19.txt 12 971 meaning File 20.txt 13 870 sentence File 21.txt 14 865 speech File 23.txt 15 849 was File 24.txt 16 832 pragmatic 17 827 will 18 823 example 19 815 use 20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
File 20.txt 13 870 sentence File 21.txt 14 865 speech File 23.txt 15 849 was File 24.txt 16 832 pragmatic 17 827 will 18 823 example 19 815 use 20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
14 865 speech File 23.txt 15 849 was 16 832 pragmatic 17 827 will 18 823 example 19 815 use 20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
File 23.txt File 23.txt File 24.txt 15 849 was 16 832 pragmatic 17 827 will 18 823 example 19 815 use 20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
File 24.txt 15 849 was 16 832 pragmatic 17 827 will 18 823 example 19 815 use 20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
17 827 will 18 823 example 19 815 use 19 815 use 20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
18 823 example 19 815 use 20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
19 815 use 20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
20 805 utterance 21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
21 782 may 22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
22 762 information 23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
23 731 like 24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
24 724 interpretation 25 705 would 26 676 do 27 675 see 28 642 been	
25 705 would 26 676 do 27 675 see 28 642 been Search Term ☑ Words □ Case □ Regex	
26 676 do 27 675 see 28 642 been Search Term ☑ Words □ Case □ Regex	
27 675 see 28 642 been Search Term ☑ Words □ Case □ Regex Hit Location	
27 675 see 28 642 been Search Term ☑ Words □ Case □ Regex Hit Location	
28 642 been Search Term ☑ Words □ Case □ Regex Hit Location	
Search Term 🗹 Words 🗌 Case 🗌 Regex Hit Location	
Search Term 🗹 Words 🗌 Case 🗌 Regex Hit Location	
	>
Advanced Search Only 0	
Start Stop Sort Lemma List Loaded	
Total No. Word List Loaded	

Figure 42. Word List with Stoplist, Prosody.

As you can read above, the words, now, are more specific to the Prosody area. Notice that the tokens and types numbers have also changed. But there's still an empty field: Lemma Word Form.

Lemma is a basic form of a word. For example, the verb GO, in English, has the following forms: go, goes, going, went, gone. But when we're talking about the verb, we use its lemma, the infinitive GO. Get an English Lemma list here : https://ldrv.ms/u/s!Aq3R_KrvKKr_qrdDTwzWyiMz3Y2bKw?e=b45IsV).

There's a way to associate all the forms of a word in the same line, using a Lemma List. Go back to the Tool Preferences menu > Wordlist > Lemma List. Load the list you've just downloaded (the list will be showed) and click APPLY.

*	Tool Preferences 🛛 🗕 🗖 🗙
Category Concordance Clusters/N-Grams Collocates Word List Keyword List	Word List Preferences Display Options Image: Construct of the state of the
	Lemma List Image: Clear Image: Treat Word List Range as Lemma List Range Word List Range O Use all words O Use specific words below Image: Output
	Add Word Add Add Words From File English Stop List - AntCc Open
	able about above according
	Apply Cancel

Figure 43. Loading the Lemma List.

You get a new list, with a Stop List and a Lemma List activated:

	Tool Preferen	ces Heip		
orpus Files	Conco	rdance Cor	ncordance Plot File View	v Clusters/N-Grams Collocates Word List Keyword List
ile 1.txt ile 2.txt	Lemma	Types: 13	298 Lemma Tok	ens: 240640 Search Hits: 0
File 3.txt	Rank	Freq	Lemma	Lemma Word Form(s)
ile 4.txt	1	18033	be	are 3321 be 3560 been 642 is 9288 was 849 were 373
ile 5.txt	2	3084	have	had 305 has 1118 have 1558 having 103
ile 6.txt	3	1892	pragmatic	pragmatic 832 pragmatics 1060
ile 7.txt ile 8.txt	4	1779	use	
ile 9.txt				use 815 used 588 uses 176 using 200
ile 10.txt	5	1656	can	can 1656
ile 11.txt	6	1365	speaker	speaker 1066 speakers 299
ile 12.txt	7	1357	language	language 1090 languages 267
ile 13.txt ile 14.txt	8	1351	sentence	sentence 870 sentences 478 sentencing 3
ile 15.txt	9	1290	example	example 823 examples 467
ile 16.txt	10	1260		context 1015 contexts 245
ile 17.txt			context	
ile 18.txt	11	1196	meaning	meaning 971 meanings 225
ile 19.txt ile 20.txt	12	1088	do	did 271 do 676 doing 93 done 48
ile 21.txt	13	1087	discourse	discourse 1060 discourses 27
ile 22.txt	14	1076	utterance	utterance 805 utterances 271
File 23.txt	15	977	semantic	semantic 505 semantics 472
File 24.txt	16	927		said 408 say 378 says 141
			say	
	17	905	act	act 482 acted 2 acting 10 acts 411
	18	902	see	saw 39 see 675 seeing 26 seen 149 sees 13
	19	870	speech	speech 865 speeches 5
	20	827	will	will 827
	21	818	case	case 482 cases 336
	22	816	interpretation	interpretation 724 interpretations 92
	23	782	may	may 782
	24	762	information	information 762
	25	759	give	gave 51 give 174 given 447 gives 51 giving 36
	26	746	like	like 731 liked 3 likes 12
	27	730	linguistic	linguistic 562 linguistics 168
	28	718	expression	expression 313 expressions 405
	20	110	expression	
	<	> < _ >	<	> <
	Search	lerm ⊠ V	Vords 🗌 Case 🗌 Reg	
				Advanced Search Only 0
	Sta	rt	Stop Sort	Lemma List 🗹 Loaded
otal No.				Word List Loaded
iles Processed	Sort by	Invert	Order	

Figure 44. Word List with Stoplist and Lemma list.

The usage of Lemma lists is not always interesting for your study. For example, in figure 44, the word ACT (905) contains: act (482), acted (2), acting (10) and acts (411). However, there's a catch: how do you know if act or acts are nouns or verbs¹³? This means that, as already explained before, you must plan very well the kind of research you're doing.

¹³ The concordance lines can tell you the difference. But you need to analyse one by one.

Estudos Descritivos e Linguística de Corpus Integrada à Prática Educativa

There's a big disadvantage of using AntConc (remember, is a free suite): you can't save your results. The only possibility, in the menu, is to click File > Save Output to Text File.

File	Global Settings	Tool Preferences	Help
0	pen File(s)	Ctrl+F	
0	pen Dir	Ctrl+D	
CI	ose Selected File(s)	
	ose All Files		
	OSC AITTICS		
CI	ear Tool		
CI	ear All Tools		
CI	ear All Tools and	Files	
Sa	ave Output to Text	t File Ctrl+S	
In	nport Settings from	m File	
E>	port Settings To I	ile	
Re	estore Default Set	tings	
Ex	dt		

Figure 45. Saving file in AntConc.

As a result, you get a file like this one:



Figure 46. File saved from AntConc's Word List.

Concordance

To obtain the concordance lines of a word, you just need to click it. For example, click the word PRAGMATIC (see figure 44). As a result, you get this screen:

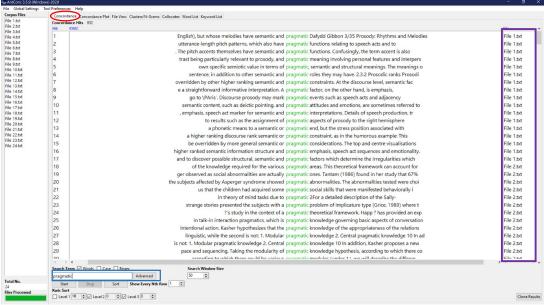


Figure 47. Concordances lines for the word PRAGMATIC.

As you can see, the program displays a list of every entry of the word PRAGMATIC (centered) in a KWIC screen (and the word appears selected as a search). On the right side, you see the name of the file the example it comes from. If you click on the word in green, you get the whole context (and the word you chose is highlighted) in the File View tab. Let's try it with the first line:

AntCone 3.5.9 (Win		
	s Tool Preferences Help	
Corpus Files	Concordance Concordance Plot File View Justers/N-Grams Collocates Word List Keyword List	
File 1.txt File 2.txt	File View Hits 17 File Tetrature in a support too simple to the property spectanae	
File 3.bt	increase on process, and may this appear too impre to use process appearance. The term 'suprasegmental' is often used as a partial synonym of 'prosody', setting the relevant features in relation to	^
File 4.bd	insection suprassignmental is solver used as a partial synonym of prosony, secting use recent results in results in results for 'segmental' units, is, e.phonemes, and implying that the features concerned are 'above' or 'larger than 'phonemes. Except for	
File 5.bd File 6.bd	seguretia units, te proteines, and importing una une reactives concerned are acover on larger unan proteines. Exception phonemic tones and phonemic pitch accents, prosociatif forms are not in great need of phonemis sair reference point,	
File 7.bd	provenic one and provenic prior access, prosonic roms are not in great need or provenies as a reference point, however, and there are independent criteria for describing and modelling them. Indeed, uterances which consist of single	
File 8.bd		
File 9.bd File 10.bd	phonemes have prosodic properties of timing and melody, so temporal size is neither a necessary nor a sufficient criterion:	
File 11.txt	"Ah!", "Oh?", "Mmm". A more congenial alternative term is 'autosegmental' (Goldsmith 1990), referring to the parallel	
File 12.txt File 13.txt	structuring of locutionary items such as syllables, words, phrases, sentences, and prosodic items. But in this tutorial the term	
File 14.txt	'prosody' is used.	
File 15.txt	The terms tone and pitch accent are used ambiguously in the literature. In the first instance they apply to the assignment	
File 16.txt File 17.txt	of a specific melody to a word to make a phonemic distinction or add a morphological function and are thus associated with	
File 18.txt	the domain of lexical prosody. The basic difference between tones and pitch accents is that a tone is one of a set of	
File 19.txt	melodies, as in Mandarin Chinese, while a pitch accent is basically a single melody which may be assigned to different	
File 20.txt File 21.txt	lexically distinctive positions in a word (as in Japanese, cf. Poser 1984). The differences are not clear-cut, however, as	
File 22.txt	Hyman (2009) has pointed out, and details differ greatly from language to language.	
File 23.bd	The term pitch accent is also often used with a different meaning, for a set of melodies which may be assigned to	
File 24.bd	different lexically distinctive stress positions in a word (as in English), but whose melodies have semantic and pragmatic	
	Dafydd Gibbon 3/35 Prosody: Rhythms and Melodies of Speech	
	rather than lexical functions, indicating old and new information, or contrast and emphasis. In addition to their association	
	with stress positions in words and phrases, they also function in larger contexts where they merge with utterance-initial or	
	utterance-final boundary tones and with overall utterance-length pitch patterns, which also have pragmatic functions relating	
	to speech acts and to dialogue management patterns such as turn-taking.	
	The English type of pitch accent is associated with stress positions in words and phrases, which are a more abstract	
	feature of language structure than their phonetic realisations as pitch accents. The stress positions in words (word stress) or	
	in phrases and sentences (phrase stress, sentence stress, nuclear stress) are the structural locations to which pitch accents	
	are assigned. The pitch accents themselves have semantic and pragmatic functions. Confusingly, the term accent is also used	
	in the literature with this abstract positional meaning, e.g. as word accent, sentence accent. The term stress is frequently	
	used in the literature with a phonetic rather than structural meaning, along with terms such as prominence and salience. The	
	definition used here are designed to provide a coherent overall terminological framework for an ontology of prosody,	
	uennuois useo nere are usagineu to provide a content overan terminological namenous no an oniciogy on prosody, which is lacking in the literature (cf. also Jassem and Gibbon 1980; Gibbon 1995; Gibbon 2006).	
	Which is lacking in the method (c), and seen in all of book in 1500, Globol 1500, G	
	the word, such as the global contour of a phrase, sentence, utterance or utterance sequence in discourse, as well as the	~
	Search Term 🖉 Words 🗌 Case 🗌 Regex Hit Location	
Total No.	pragmatic Advanced 1 0	
24 Files Processed	Start Stop	
		Clone Results

Figure 48. File View for the first entry of the word PRAGMATIC.

If you want, you can copy the text using the control + c keys. Another possibility of viewing the behavior of the word PRAGMATIC is by clicking the tab Concordance Plot:

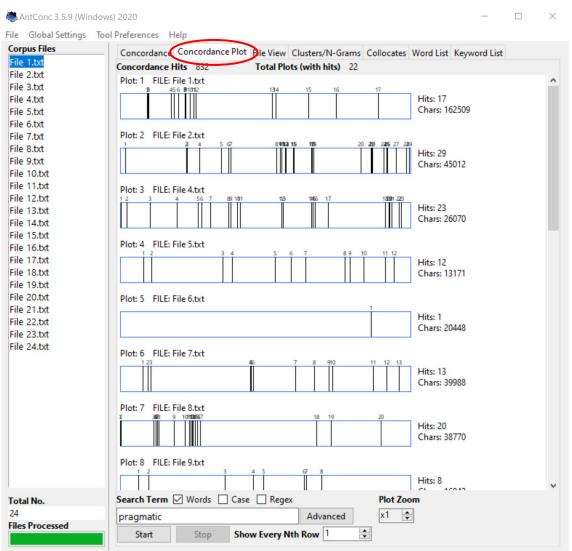


Figure 49. Concordance Plot for the word PRAGMATIC.

The slashes you see in the rectangles represent the distribution of the word PRAGMATIC in each text of the corpus.

PRAGMATIC, as you noticed, is an adjective. The name of the area is PRAGMATICS (a noun). Let's search it? You just need to go back to the Concordance tab and write the word in the search field (as in figure 47) and click START. Then we get:

AntConc 3.5.9 (Windo			- 0
	Tool Preferences Help		
rpus Files		ew Clusters/N-Grams Collocates Word List Keyword List	
2.txt	Concordance Hits 1050		File
e 3.txt	Hit KWIC		
4.bd	1	rse semantics, such as information structure, and pragmatics, such as speaker or hearer orientation, rather	File 1.txt
e 5.bd	2	1 Autism and Pragmatics of Language Asa Kasher1 and Sara Meilijson	File 2.txt
7.bd	3	ndrome using the theoretical framework of Modular Pragmatics of Language. We review definitions and classifica	File 2.bd
8.bd	4	this paper the point of view of Pragmatics of Language, based on the theoretical model	File 2.txt
9.bd 10.bd	5	that in Autism, theory of mind and pragmatics of language are impaired, while syntax and	File 2.txt
11.txt	6	. However, most references to the impairments in pragmatics of language in the literature are not	File 2.txt
12.bd	7	we present a theoretical framework of Modular Pragmatics of Language that can account for the	File 2.txt
13.bd 14.bd	8	skills with abnormalities in the area of pragmatics. Lack of nonverbal expressiveness, impression of	File 2.txt
15.bd	9	. While it is now widely accepted that pragmatics is the most disturbed in Autism, it	File 2.txt
16.txt 17.txt	10	f abnormal rather than normal development. 3. The Pragmatics of Language point of view Research in	File 2.txt
17.txt 18.txt	11	in Autism lies in the area of pragmatics of language and propose to approach the	File 2.txt
19.txt	12	with confusing results. Lately, the aspect of pragmatics has become prominent in research and it	File 2.txt
20.txt 21.txt	13	accepted that difficulties in the domain of pragmatics are present in Autism (Rapin, 1991). However, mos	File 2.txt
22.txt	14	the impairment lies in the area of pragmatics of language and not in communication as	File 2.txt
3.txt	15	here is based on the view of pragmatics as a study of language use that	File 2.bt
24.txt	16	distinguishes between language and communication. Pragmatics is viewed as the linguistic conditions of	File 2.txt
	17	to be included in talk-in interaction pragmatics, which is pragmatic knowledge governing basic asp	File 2.bt
	18	with each other under the heading of "pragmatics". In his view, a distinction should be	File 2.txt
	19	are manifested in the autistic individual. Core Pragmatics. The knowledge of basic speech act types	File 2.txt
	20	ted solely by environmental needs. Amplified Core Pragmatics. The knowledge of all the systems of	File 2.txt
	21	to do with words. Talk-in Interaction Pragmatics. The pragmatic knowledge governing basic aspects	File 2.txt
	22	rcio & Paccia 1987). Deficiencies in this area of pragmatics have a great influence in the social	File 2.txt
	23	describes her difficulty in this area of pragmatics as a lack of the concept of "	File 2.bxt
	24	a conversation for company's sake. Central Pragmatics. Knowledge related to general cognitive systems a	File 2.txt
	25	ships with other people (Hobson, 1989). Interface Pragmatics. The knowledge which involves integration of data	File 2.txt
	26	deficient in autistic individuals. This area of pragmatics includes all paralinguistic and nonverbal aspects	File 2.txt
	27	space" (p. 70). Difficulties in this area of pragmatics are beautifully depicted by Williams when she	File 2.txt
	28	ic distribution). 14 The theoretical framework of pragmatics of language as described above can account	File 2.txt
	29	the social interaction area pertain to Interface Pragmatics, "marked impairment in the use of multiple	File 2.txt
	20	in Communication area notation to definite in accomption of language. The first one "delay in	Eilo 2 tot
	Search Term 🗹 Words 🗌 Case 🗌 R	earch Window Size	
		egec Search Window Size	
I No.	pragmatics		
	Start Stop Sort Kwic Sort	Show Every Nth Row 1 2	
Processed	Level 1 1R C Level 20 0		Clone R

Figure 50. Choosing the word PRAGMATICS for concordance lines.

As you can see in figure 50, there are **parameters** you can change to show different results. Let's try with this configuration: Vevel 1 R Configuration (remember: choose your configurations and click the START button). The result we get is this:

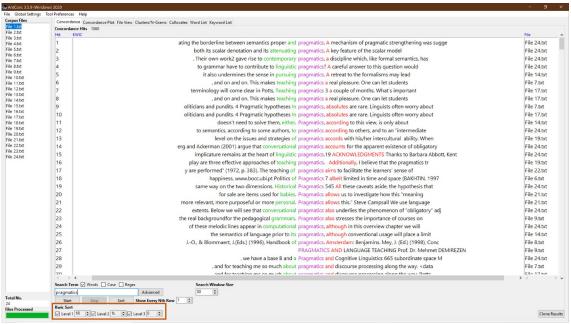


Figure 51. Different levels of sorting with PRAGMATICS concordance lines.

The parameters we choose must reflect our research. In this case, we're just playing with the tool, but for a specific research you need to choose specific parameters.

Yet another possibility of analyzing the word you are interested in is to search the words that co-occur with it. Go back to Tools Preferences and erase the configurations we did for the StopList and LemmaList, and then load the WordList again (with no search term); after that, write the word SPEAKER in the same field and click the button search only (figure 52).

e Global Settings	Tool Preferen	ces Help				
orpus Files	Conco	rdance Co	incordance Plot File View Cluste	ers/N-Grams Collocates Word List Key	word List	
ile 1.txt		ypes: 173				
ile 2.txt	Rank	Freq	Word	Lemma Word Form(s)		-
le 3.txt le 4.txt	30	1582	wnat			
le 5.txt	37	1558	have			
le 6.txt	38	1410	one			
e 7.txt	39	1373	I.			
e 8.txt	40	1371				
le 9.txt			they			
e 10.txt e 11.txt	41	1306	at			
le 12.txt	42	1296	f			
e 13.txt	43	1285	m			
le 14.txt	44	1251	there			
e 15.txt	45		d			
le 16.txt		1248	a			
le 17.txt le 18.txt	46	1164	р			
e 19.txt	47	1118	has			
le 20.txt	48	1090	language			
le 21.txt	49	1066	speaker			
le 22.txt	50	1060	discourse			
le 23.txt le 24.txt						
16 24.LAL	51	1060	pragmatics			
	52	1043	these			
	53	1032	other			
	54	1015	context			
	55	981	such			
	56	980	some			
	57	972	if			
	58	971	meaning			
	59	955	g			
	60	938	he			
		100 A 100 A				
	61	928	more			
	<	> < >	<	> <	3	> •
	Search	Term 🗹	Words 🗌 Case 🗌 Regex	Hit Location		
	speake	er	А	dvanced Search Only 1		
				Lemma List Loaded		
tal No.	Sta	art	Stop Sort	Word List Loaded		
4	Cort hu	Inver	Order			
es Processed	Sort by		t ofder			

Figure 52. Searching for a specific word in the WordList.

To continue with the analysis of the word SPEAKER, now we're going to work with the tab Collocates. Just choose the tab, certify that the word SPEAKER is still in the search field and click Start. The result you can see in igure 53 (in this case, we used the sort by frequency option).

le Global Settings	ndows) 2020	cor List-				- 0
Corpus Files						
File 1.txt	-					s/N-Grams Collocates Word List Keyword List
ile 2.txt		o. of Colloo				lo. of Collocate Tokens: 10660
ile 3.txt	Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
ile 4.txt	1	1230	923	307	4.33745	the
ile 5.txt	2	377	168	209	3.96391	to
ile 6.txt ile 7.txt	3	349	31	318	4.92334	s
ile 8.txt	4	348	230	118	3.17476	of
ile 9.txt	5			129		
ile 10.txt		343	214		3.38200	a
ile 11.txt	6	266	182	84	3.99527	that
ile 12.txt ile 13.txt	7	260	111	149	3.63996	is
ile 14.txt	8	248	83	165	3.31117	and
ile 15.txt	9	196	94	102	3.03137	in
ile 16.txt	10	101	74	27	4.82942	what
ile 17.txt ile 18.txt	11	100	70	30	4.11001	by
File 19.txt	12	98	38	60	3.97363	or
File 20.txt	13	94	27	67	6.89801	hearer
ile 21.txt ile 22.txt						it
ile 23.txt	14	79	30	49	3.40382	
ile 24.txt	15	75	29	46	3.49911	not
	16	74	38	36	3.59193	with
	17	74	17	57	7.29395	addressee
	18	73	29	44	2.91946	as
	19	72	40	32	3.06141	for
	20	69	22	47	3.10961	be
	21	61	38	23	3.50067	on
	22	60	12	48	4.78231	meaning
	23	58	36	22	4.94937	about
	24	52	14	38	4.37248	has
	25	50	21	29	2.83898	this
	26	47	29	18	3.30349	which
	20	47	29	10	3.50549	
	< 2	> < >	< >	< >		
	Search	Term 🗹	Words	Case	Regex	Window Span 🗹 Same
	speake	er			Adv	vanced From 5L 束 To 5R 🌩
otal No.	Sta	irt	Stop	Sor	t	Min. Collocate Frequency
iles Processed	Sort by	Invert	t Order			1
nes rrocessed	Sort by	Freq	~			Clone Resul

Figure 53. Collocates for the word SPEAKER.

Observe the example of the word HEARER: it occurs 94 times in the corpus, 27 times to the left of HEARER and 67 times to the right of it. If you click the word HEARER in this screen, you go back to the Concordance tab with these combinations and you get

Lexical Analysis Tools

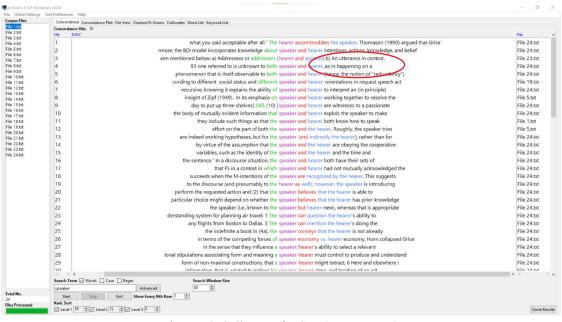


Figure 54. Collocates for SPEAKER + HEARER.

The final tab related to the Concordance tool is the Clusters/N-Grams. Still using the same screen, click on the tab CLUSTERS/N-Grams and the START button:

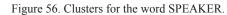
· ·	dows) 2020				
le Global Settings	Tool Preferen	ces Help			
orpus Files	Conco	rdance Co	oncordanc	e Plot File View Clusters/N-Grams Collocates Word List Keyword List	
ile 1.txt	Total N	o. of Clust	er Types	2989 Total No. of Cluster Tokens 4264	
ile 2.txt ile 3.txt	Rank	Freq	Range	Cluster	
ile 4.txt	1	255	12	speaker's	
ile 5.txt	2	56	6	speaker is	
ile 6.txt	3	49	7		
ile 7.txt		0.00	-	speaker and	
ile 8.txt ile 9.txt	4	32	7	speaker to	
ile 10.txt	5	30	6	speaker has	
ile 11.txt	6	25	2	speaker's meaning	
ile 12.txt	7	23	1	speaker means	
ile 13.txt	8	21	2	speaker's intention	
ile 14.txt		-	-	• 1000 (1000) (1000) (1000)	
ile 15.txt ile 16.txt	9	18	4	speaker of	
ile 17.txt	10	16	1	speaker meaning	
ile 18.txt	11	15	1	speaker and addressee	
ile 19.txt	12	14	2	speaker who	
ile 20.txt	13	14	1	speaker's empathy	
ile 21.txt ile 22.txt					
ile 23.txt	14	13	2	speaker intends	
ile 24.txt	15	13	1	speaker says	
	16	12	3	speaker may	
	17	12	5	speaker or	
	18	11	4	speaker and hearer	
	19	11	2	speaker intends to	
		2.0	1.1.1	•	
	20	11	3	speaker must	
	21	11	1	speaker's empathy with	
	22	11	1	speaker's intentions	
	23	11	4	speaker's utterance	
	24	10	1	speaker does	
	25	10	3	speaker intentions	
	26	10	2	speaker wants	
	27	10	2	speaker's communicative	
	28	10	3	speaker's intended	
	< .	> < >	- < - 3	<	>
	Search	Term 🗹	Words	Case Regex N-Grams Cluster Size	
	speake	er		Advanced Min. 2 🗭 Max. 5 🚔	
	Sta	rt	Stop	Sort Min. Freq. Min. Range	
otal No.		Invert		Search Term Position	
4	Sort by			✓ On Left On Right	
les Processed	Sort by	rieq	~		

Figure 55. Clusters for the word SPEAKER.

In this case, we have a cluster size with 2 to 5 (from 2 to 5 words). We can change it, for example, to 2 to 4:

Lexical Analysis Tools

L. CLARKER	THE				
le Global Settings					
Corpus Files File 1.txt				e Plot File View Clusters/N-Grams Collocates Word List Keyword List	
File 2.txt		o. of Clust			
ile 3.txt	Rank	Freq	Range	Cluster	
ile 4.txt	1	255	12	speaker's	
ile 5.txt	2	56	6	speaker is	
ile 6.txt ile 7.txt	3	49	7	speaker and	
ile 8.txt	4	32	7	speaker to	
le 9.txt	5				
le 10.txt		30	6	speaker has	
le 11.txt	6	25	2	speaker's meaning	
le 12.txt le 13.txt	7	23	1	speaker means	
le 14.txt	8	21	2	speaker's intention	
le 15.txt	9	18	4	speaker of	
le 16.txt	10	16	1	speaker meaning	
ile 17.txt ile 18.txt	11	15	1	speaker and addressee	
ile 19.txt	12	14	2		
ile 20.txt				speaker who	
ile 21.txt	13	14	1	speaker's empathy	
ile 22.txt ile 23.txt	14	13	2	speaker intends	
ile 24.txt	15	13	1	speaker says	
	16	12	3	speaker may	
	17	12	5	speaker or	
	18	11	4	speaker and hearer	
	19	11	2	speaker intends to	
	20	11	3	speaker must	
			-		
	21	11	1	speaker's empathy with	
	22	11	1	speaker's intentions	
	23	11	4	speaker's utterance	
	24	10	1	speaker does	
	25	10	3	speaker intentions	
	26	10	2	speaker wants	
	27	10	2	speaker's communicative	
	28	10	3	speaker's intended	
	20	10	P	speaker's interfueu	
	S	Term I	Mard-	Case Regex N-Grams Cluster Size	2
			words		
	speake			Advanced Min. 2 Adva. 4	
	Sta		Stop	Sort Min. Freq. Min. Range	
otal No.	Sort by	Inver	t Order	Search Term Position 1 🖨 1	



We can search the cluster SPEAKER'S UTTERANCE, for example (note that we changed the Search Window Size parameters to 100, to visualize larger fragments of text):

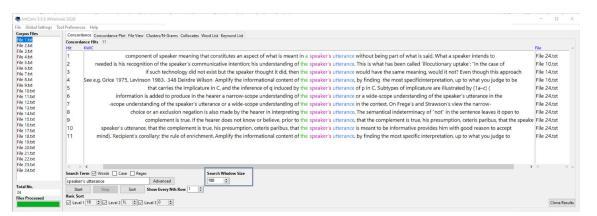


Figure 57. Concordance lines for the cluster SPEAKER'S UTTERANCE.

Remember, the lexical analysis programs exist to help, not to do the entire work for you. In figure 57, for example, you can start analyzing the SPEAKER'S UTTERANCE cluster for your research.

Keywords

To find the keywords of a corpus, you need to have two available corpora: your study corpus and a reference corpus. You can download a reference corpus list from here: https://ldrv.ms/u/s!Aq3R_KrvKKr_qrdDTwzWyiMz3Y2bKw?e=b45IsV

The first step is to load your reference corpus. Go to the upper menu of AntConc again and choose Tool Preferences > Keyword List > Reference Corpus > Use Word List > Add File. Get the file you've just downloaded and click load. The name of the corpus will appear in the rectangle. After that, click Load. The program will load the reference corpus and then will show a green rectangle, showing that everything is ok. Click APPLY to go

Lexical Analysis Tools

Stool Preferences		_		×
Category Concordance Concordance Plot Clusters/N-Grams Collocates Word List Keyword List	Keyword List Preferences Display Options			
		Apply	Can	cel

Figure 58. AntConc's Tool Preferences and Keyword List.

You have already changed the settings and loaded the reference corpus. Just choose the Keyword List tab of AntConc and click the START button. You get your keywords list from the PRAGMATICS area:

ile Global Settings	Tool Prefere	nces Heln				
Corpus Files					No. I No. I No. I No. I Konserved	17.4
File 1.txt	1			e View Clusters/ ord Tokens: 2	N-Grams Collocates Word List Keyword 54006 Search Hits: 0	LIST
File 2.txt	Rank	rd Types: Freq	Keyness	Effect	Keyword	
File 3.txt	1	1060	+ 13792.5	0.0088		
File 4.txt File 5.txt					pragmatics	
ile 6.txt	2	1656	+ 13327.94	0.0132	can	
ile 7.txt	3	870	+ 11074.27	0.0072	sentence	
ile 8.txt	4	805	+ 8965.74	0.0066	utterance	
ile 9.txt ile 10.txt	5	971	+ 7978.52	0.0079	meaning	
ile 11.txt	6	827	+ 7908.47	0.0068	will	
ile 12.txt	7	1060	+ 7522.3	0.0084	discourse	
ile 13.txt	8	731	+ 7447.96	0.006	like	
ile 14.txt ile 15.txt	9	832	+ 7431.83	0.0068	1000000	
ile 16.txt	-				pragmatic	
ile 17.txt	10	1066	+ 6925.03	0.0083	speaker	
ile 18.txt	11	478	+ 6267.82	0.004	sentences	
ile 19.txt ile 20.txt	12	421	+ 5768.35	0.0035	john	
ile 21.txt	13	425	+ 5456.1	0.0035	new	
ile 22.txt	14	870	+ 5336.89	0.0068	sentence	
ile 23.txt	15	482	+ 5208.62	0.004	case	
ile 24.txt	16	1015	+ 5112.34	0.0075	context	
	17	474	+ 5053.77	0.0039	content	
	18	472	+ 4954.71	0.0039	semantics	
	19	971	+ 4895.23	0.0072	meaning	
	20	349	+ 4781.73	0.0029	grice	
	21	370	+ 4698.95	0.0031	words	
	22	405	+ 4674.79	0.0034	structure	
	23	364	+ 4462.35	0.003	even	
	24	505	+ 4445.39	0.0042	semantic	
	25	420	+ 4358.09	0.0035	reference	
	26	562	+ 4314.39	0.0046	linguistic	
	27	724	+ 4279.76	0.0057	-	
	21	124	+ 42/9./0	0.0057	interpretation	
	<	> < _>	< >	<	> <	> •
	Search	Term 🗹 \	Vords 🗌 Case 🗌		Hit Location	
	speak	er		Adva	enced Search Only 0	
otal No.	St	art	Stop So	rt	Reference Corpus 🛛 Loaded	
iles Processed	Sort by	Invert	Order			
lies Processed	Sort b	/ Keyness	~			Clone Results

Figure 59. Keyword List from the Prosody corpus.

From this point on, you already know the procedures: click the word you want to analyze and you get its concordance lines. For example, let's click the word DISCOURSE (rank 7) and see the results:

Lexical Analysis Tools

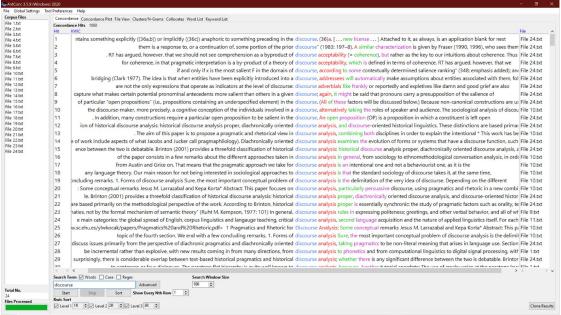


Figure 60. Concordance lines for the word DISCOURSE

From here, you can try the tabs we have already explained, like Concordance Plot, Collocates and Clusters/N-Grams. Have fun!

Project Development

Basic Contents

- Describing language nowadays.
- Choosing an area to be described: Terminography.
- Web environment for terminological management: VoTec.

Objectives

- Recognize the steps of a language description project.
- Prepare a project you can use with your students.

PROJECT DEVELOPMENT

ACTIVITY 11

Video class, module 4. Watch the professor's hints about the subjects that are going to be worked in this module.

ACTIVITY 12

Project. You're going to prepare a class plan project, using the Corpus Linguistics approach. Possible sources: Internet sites with corpora and tools (module 2), corpus compilation (module 3), lexical analysis tools software (module 3), a combination of some or all this previous ideas. Here in module 4 we present a possible project you could work with your students. Think about a project you really could apply in your school.

The project must contain all the steps you intend to develop with the students (don't forget to specify the target public, the description of the school you plan to apply the project, the available resources in this school, etc.). You've seen in this guide that pictures help a lot to prepare this kind of work – use them. When you finished it, send to your tutor. This activity will be evaluated.

Describing language nowadays

There are three main sources from where you can get information about a language: dictionaries, didactic materials and grammar books (when we think about the teaching/ learning relation in schools, of course). Nevertheless, it's a very passive action: you open the books and consult them; if you don't understand what is written or you don't get the information you need, you need the help of a teacher or get another material.

With the Corpus Linguistics approach, these classical learning methods are being reviewed. It's time to teach your students how and where they can get information about the language in a more active way! It's time to teach them how to solve their own questions and go beyond without the help of a teacher.

Among other areas, the Corpus Linguistics approach can work with: Phonetics and Phonology; frequency of the most common words in a language; frequency of grammatical classes; Morphology and morphossyntatic variations; Syntax; collocation comparisons in a language (such as adverbials, for example); recognition of composite (as binomials) and complex (or n- grams) Lexicon; Phraseologisms (sayings, idioms); verbal and nominal regency; selection of nomenclature and database forterminological works (in all areas); general mono-or multilingual dictionaries creation; verification of translation modalities in mono-or bilingual corpora; Dialectology; translatorsdatabase; foreign language teaching and development of teaching materials (ESL); literary, technical and journalistic (parallel corpora) translations review; aid for Text, Gender and Discourse studies; Prosody; creation of computerized tools (such as lexical and grammar checkers, summarization, text mining, textual simplification and elaboration); natural language processing (NLP) - automated translations; Interlanguage description in foreign language teaching; Stylistics analysis; Pragmatics; description and analysis of "errors" in written texts in native or foreign language; Semantics; human and machine translation; Metaphors study and treatment.

There are many possibilities of using Corpus Linguistics in other areas than Linguistics and Language. Learn about one here, with the text of Oliria Mendes Gimenes: https://repositorio.ufu.br/handle/123456789/13884

Choosing an area to be described: Terminography

Terminology is the branch of Linguistics that studies the vocabulary of a specific area of knowledge. For example, we can study the terminology of Chemistry, the terminology of Education, etc. Terminography is the practical branch related to Terminology. When you create a specialized dictionary for an area of knowledge, you are doing a terminographical work.

Terminographical works generally, nowadays, use corpora as basis to create the microstructure of a dictionary entry. The examples we get from a corpus help us to determine the grammar structure of the word, its field, its semantical relations, etc. More important, they help us build the definition of a specialized word.

To obtain the examples, just access the corpus, using the AntConc program (as we have previously done), and go to the concordance lines: they supply information you can use in the dictionary's microstructure.

For the definition, a good idea is to seek the word you want plus the verb to be. Take a look at figure 62. It's the same of figure 61, but we added the word *is* in the search field:

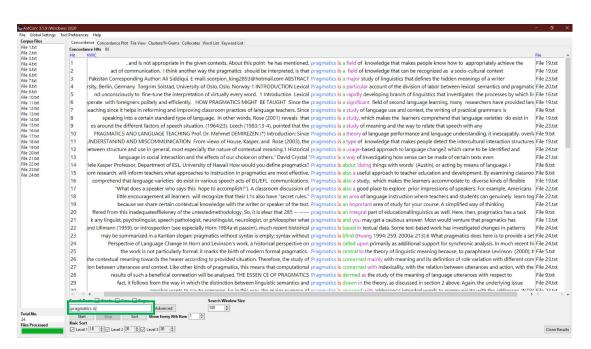


Figure 61. Concordances for the combination PRAGMATICS + IS.

The concordance lines you get are serious candidates of what Aubert (1996) calls Definitory Contexts, a good definition of the term. In this case, we have 80 concordance lines that can supply us with information about what is PRAGMATICS. Other possibilities of search to find Definitory Contexts are apposition (using commas), brackets and colon In figure 62, we searched PRAGMATICS + colon:

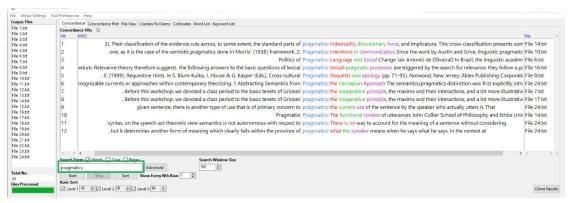


Figure 62. Concordance for the combination PRAGMATICS + :..

To organize all the information you get from the concordance lines, in terminographical works, you need a specific program. The program can be installed in your computer (for example, some researchers still use Access, from Microsoft) or you can use online solutions.

Web environment for terminological management: VoTec

The VoTec (Vocabulário Técnico; FROMM, 2007) is an online solution for terminographical projects. Nowadays we call this type of program web environment for terminological management. You can see some examples of how the system works in here: http://treino.votec.ileel.ufu.br/. This is the page of the product of researches done with the VoTec¹⁴.

Online Technical Vocabu	lary	Full Screen Português Help
Linguistics V C	hoose an area	
Kind of exhibition Normal Extended Kind of search Total Translator Modular External Searches Corpus NILC Google Answers.com Wikipedia CORTEC	▼ English Allophone in the area: Linguistics > Theoretical Linguistics > Phonetics Derivation in the area: Linguistics > Theoretical Linguistics > Morphology, Grammar in the area: Linguistics > Theoretical Linguistics > Syntax Intonation in the area: Linguistics > Theoretical Linguistics > Syntax Intonation in the area: Linguistics > Theoretical Linguistics > Syntax Magning in the area: Linguistics > Theoretical Linguistics > Syntax Maning in the area: Linguistics > Theoretical Linguistics > Syntax Maning in the area: Linguistics > Theoretical Linguistics > Syntax Morphology in the area: Linguistics > Theoretical Linguistics > Morphology Morphology in the area: Linguistics > Theoretical Linguistics > Morphology Phoneme in the area: Linguistics > Theoretical Linguistics > Morphology Phonetics in the area: Linguistics > Theoretical Linguistics > Morphology Phonetics in the area: Linguistics > Theoretical Linguistics > Phonology Phonetics in the area: Linguistics > Theoretical Linguistics > Phonology Phonetics in the area: Linguistics > Theoretical Linguistics > Phonology Phonetics in the area: Linguistics > Theoretical Linguistics > Phonology Speech in the area: Linguistics > Theoretical Linguistics > Phonology Stem in the area: Linguistics > Theoretical Linguistics > Phonology Symphonetics in the area: Linguistics > Theoretical Linguistics > Phonology Speech in the area: Linguistics > Theoretical Linguistics > Morphology Vowels in the area: Linguistics > Theoretical Linguistics > Phonology Speech in the area: Linguistics > Theoretical Linguistics > Phonology Symphonetics in the area: Linguistics > Theoretical Linguistics > Phonology Speech in the area: Linguistics > Theoretical Linguistics > Phonology Symphonetics in the area: Linguistics > Theoretical Linguistics > Phonology Speech in the area: Linguistics > Theoretical Linguist	
	 Português Afixos na área: Linguística > Linguística Teórica > Morfologia Alofone na área: Linguística > Linguística Teórica > Morfologia Alofone na área: Linguística > Linguística Teórica > Morfologia Entonação na área: Linguística > Linguística Teórica > Fonedica Portvação na área: Linguística > Linguística Teórica > Fonologia Fonema na área: Linguística > Linguística Teórica > Fonologia Fonema na área: Linguística > Linguística Teórica > Fonologia Fonologia na área: Linguística > Linguística Teórica > Fonologia Fonologia na área: Linguística > Linguística Teórica > Fonologia Fonologia na área: Linguística > Linguística Teórica > Fonologia Fonótica na área: Linguística > Linguística Teórica > Fonética Fonética na área: Linguística > Linguística Teórica > Fonética Gramática na área: Linguística > Linguística Teórica > Fonética Gramática na área: Linguística > Linguística Teórica > Sintaxe Línguística > Linguística Teórica > Morfologia Radical na área: Linguística > Linguística Teórica > Morfologia Significado na área: Linguística > Linguística Teórica > Sintaxe UF na área: Linguística > Linguística Teórica > Sintaxe UF na área: Linguística > Linguística Teórica > Sintaxe Yogais na área: Linguística > Linguística Teórica > Sintaxe Yogais na área: Linguística > Linguística Teórica > Sintaxe 	

Figure 63. Main consulting page of VoTec (in English), Linguistics area.

VoTec was developed as a bilingual (English/Portuguese) tool. If you develop a project, you must work with these two languages. As a result, all the processes we presented

The first step in VoTec is to have your name registered in the system. As soon as you're granted access to the system, you can start creating a term.

¹⁴ Take a look at new possibilities with this version of VoTec, that works with TV Series: http://ic.votec. ileel.ufu.br

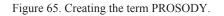
If you're interested in using VoTec for the class plan project of activity 12, get in touch with your tutor. We must register your name in the system first so you can access the data bank. In this case, you're going to work with the pages of the **process** involved in VoTec to create a specialized dictionary.

Let's see the first page:

Vocabulário Te	écnico Online			Tela Cheia English
Bem Vin	do, tester			
Fichas aind	Troc	Seu último ac ar Senha Novo Termo das	esso foi em: Cadastro de Fontes Sair	
	-			
Termo	Língua	Coletado em	Cadastrada por	Ações
		Coletado em	Cadastrada por	Ações
Termo Nenhum termo pa		Coletado em	Cadastrada por	Ações
Termo Nenhum termo pa Total de Termos: 0			<u>Cadastrada por</u>	Ações
Termo Nenhum termo pa Total de Termos: 0	ara revisar		Cadastrada por	Ações Ações

The first step is to create a NEW TERM. Click the button; insert the name of the term and the areas involved¹⁵:

Vocabulário Técnico Online Novo Termo	Tela Cheia English
Voltar ao painel	
Passo 1	
Termo	
Prosody	
Língua	
Escolha uma língua: Inglês 🗸	
Ontologia	
Grande Área: Linguística 🗸	
Sub-área 1: Linguística Teórica 🗸 🗸	
Sub-área 2: Prosódia 🗸	~



¹⁵ Here, instead of the Pragmatics area, we're working with the Prosody area.

Figure 64. VoTec researcher's first page.

In the next screen, you get the name of the term you're going to work with and you can start inserting the concordance lines (contexts) you got from AntConc. In this case, let's insert the information from the fourth concordance line for PROSDOY + IS (figure 66):

/ocabulário Técnico Onlin	e		Tela Cheia English
Cadastro de Cor	ntextos para Proso	ody	
	Voltar ao Painel	Próximo Passo	
Novo Contexto	L .		
Passo 2			
Dados do Contexto			
hall	sodic reading, or reading with expre imarks of fluent reading. The major ding prosody is related to decoding a	purpose of the study was to learn	
Conceito*: pro	sody is related with expression		
Fonte*: OPE	DF		✓ Cadastrar Nova
Data de Coleta*: 20,	(dia/mês/a	ino ex.: 18/03/2007)	~
Contextos Cadastrad	os		
Exemplo	Conceito	Fonte	Ações
Nenhum contexto cadastrado			
Contextos Cadastrados: 0			
			14 06:53 © 2007 FFLCH - ICMC Jr.

Figure 66. Inserting a context for PROSODY.

As you can see, after inserting an extract of a concordance line from AntConc, we must present a concept for it (summarize the main idea). The font can continue 0PDF and you must insert a date. After you do these steps, click the **SAVE** button.

You don't have to insert all the contexts, just the ones you consider the best to create a definition. It means you must understand the text very well and you must be capable of extracting some basic concepts from it. These basic concepts will form your definition.

As a result, we get a new screen, with the context already inserted in the data bank. In example 56, we got a screen with 3 examples

Project Development

Exemplo	Conceito	Fonte	Ações
Although these studies provide good descriptive evidence for the idea that prosody is linked to fluent and expressive reading, they do not help us understand whether prosody is helpful for comprehension in Young readers	linked to fluent and expressive reading .	PDF 20/04/2014	editar - excluir
Prosodic reading, or reading with expression, is considered one of the hallmarks of fluent reading. The major purpose of the study was to learn how reading prosody is related to decoding and Reading comprehension skills.	prosody is related with expression	PDF 20/04/2014	editar - excluir
As noted earlier, Chafe (1988) and others suggested that prosody is a reflection of comprehension.	reflection of comprehension	PDF 20/04/2014	editar - excluir

Figure 67. Inserted contexts for PROSODY.

When you finish inserting the contexts, click the button **NEXT STEP** (figure 66). You get a new page with this appearance:

<u> </u>	rmo: Prosody				
	textos Conceito Final / Definição	asso Anter	ior Salvar Sair sem salvar		
201	Exemplo		Conceito	Fonte	
1	Prosodic reading, or reading with ex is considered one of the hallmarks or reading. The major purpose of the s to learn how reading prosody is rela decoding and Reading comprehension	of fluent study was ted to	prosody is related with expression	PDF 20/04/2014	
2	As noted earlier, Chafe (1988) and c suggested that prosody is a reflectio comprehension.		reflection of comprehension	PDF 20/04/2014	
3 Although these studies provide good descriptive evidence for the idea that prosody linked to fluent and expressive reading PDF 20/04/2014					
	os / Traços Distintivos / Semânti dos-	ca Tern	no Equivalente 🛛 Termos Remissivos 🖌 Info	ormações Enciclopédic	as
	Ontologia: Linguística > L	inguística T	eórica > Prosódia	\bigcirc	^
(Categoria Gramatical: Escolha uma (Categoria 🗸	Número: Singular 🗸		
	Gênero: Masculino 🗸		Sigla/Acrônimo:		
	Entrada por Extenso:				
	Var. Morfossintáticas:				
	Acepção Nº:				~

Figure 68. Second page for the term PROSODY

Each of the tabs in the middle of the page must be completed with the information from the concordance lines you have already inserted in (that is shown in the first part of the page, as you can see in figure 68). Let's see one by one:

The DATA tab must be filled with morphossyntatic and corpus information about the term you're working with:

Ontologia:	Linguística > Linguística Teórica > Prosódia	^
Categoria Gramatical:	Substantivo V Número: Singular V	·
Gênero:	Neutro Y Sigla/Acrônimo:	
Entrada por Extenso:		
Var. Morfossintáticas:		
Acepção Nº:		~
Córpus		
Posição na Ordem de Fregüência:	4 N° de Ocorrências do 388 termo:	、

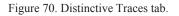
Figure 69. Data tab.

To get the position in frequency order and the number of occurrences of the term, get back to the AntConc's Word List.

Almost hundred percent of words used in technical areas are nouns. Moreover, almost hundred percent of the English nouns have a neutral gender.

The Distinctive Traces tab must contain the concepts you prepared in the previous page. The lines represent the number of contexts you inserted in the system, the columns, the synonyms you can get from these concepts. As you can see in figure 70, we inserted the concepts expression and expressive reading in the same column (A):

-	os Traços D ços Distintivos-	istintivos y Se	emantica	Termo Equiv	/alente / T	ermos Remissivos	Informações Enciclopédicas
Nc	ova Coluna 📃	Não use aspas d	uplas				
_			-	-	_		
	Α	P	C	D	E		
1	expression						
2		comprehens					
3	expressive		fluency				
_	<u>ь</u> е						



In the Semantics tab, you must get dictionary information and all the semantic traces you can get from the examples:

Project Development

Termo Dicionarizado?	✓ Definições Coincidentes? Parcial ✓		^
Fonte:	Merriam-Webster]	
Definição Dicionarizada:	the rhythm and pattern of sounds of poetry and language	^	L
		\sim	
Hiperônimo de:			
Hipônimo de:			~
Co-hipônimo de:	comprehension	×	
Sinônimo(s):	Antônimo(s):		ł.
Notas:		~	
		\checkmark	~

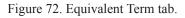
Figure 71. Semantics tab.

Don't worry about the definition you copy from other dictionary (DICTIONARIZED INFORMATION). They'll not be shown in the main consulting page of VoTec. We fill in the field just to help us understand the term better. In the NOTES field, you can write anything you consider important for you (also not shown in the consulting page).

Doubts about the terms you read in figure 71? Search them in the Linguistics area (see picture 54) of VoTec (you'll find the explanations in Portuguese).

We have already explained that VoTec is a bilingual platform. As you as you have inserted the terms in both languages (in this case: PROSODY and PROSÓDIA), you must link hem through the EQUIVALENT TERM tab:

Dados Traços Distintivo Termo Equivalente	os Semântica Termo Equivalente T ermos Remissivos Informações Enciclopédicas	
	O termo equivalente deve ser um termo cadastrado em inglês, com a mesma ontologia e que ainda não esteja marcado como equivalente de outro termo.	
Buscar Termo Equivalente:	Buscar	



In the CROSS-REFERENCE TERMS tab, if available (it means that this term must already been inserted in the system), you can link the termyou're working with terms that you consider relevant for a better comprehension of it.

Dados / Traços Distintivos - Termos Remissivos	s Semântica Termo Equivalente Termos Remissivos finformações Enciclopédic	as
	Os termos remissivos devem ser termos cadastrado em português e com os respectivos cada já aprovados pelo administrador.	astro
Buscar Termo Remissivo:	Buscar	
Termo	Ap	oagar

Figure 73. Cross-reference terms tab.

The information you insert in the ENCYCLOPEDIC INFORMATION tab can come from

Wikipedia:

Definição:					^
Artigo:					
Fonte:					
Link:		Tipo	: Imagem 🗸	·	
	Adicionar				

Figure 74. Encyclopedic Information tab.

The final step, using the DISTINCTIVE TRACES tab, is to fill in the FINAL CONCEPT and DEFINITION fields.

Voca	abulário Téci	nico Online					Tela Cheia Engli
Те	rmo: Pr	osody					
			Passo A	Interior	Salvar Sa	air sem salvar	
	ntextos Cono nceito Final / D	ceito Final / Def	inição				
		,					
	Conceito	Final: deals w	ith language	expression,	comprehensio	on and fluency	^
							\sim
	Def					1.0	
	Der	inição: <u>deals w</u>	ith language	expression,	comprehensio	on and fluency	~
							~
		L					
Dac		vistintivos Se	mântica	Termo Equi	valente T	ermos Remissivos	Informações Enciclopédicas
	ços Distintivos						
N	ova Coluna 🗕	Não use aspas d	uplas				
	A	В	с	D	E]	
1	expression]	
2		comprehens				1	

Figure 75. Final Concept and Definition tabs.

Project Development

The FINAL CONCEPT tab is used to join the concepts you have in the DISTINCTIVE TRACES tab. As a definition is a sentence, you must rewrite them as a clear and synthetic sentence. After you finish these processes and the term is approved, you got a screen like this one:

Vocabulário Técnico Onlin	e Tela Cheia English Ajuda
Linguística V Escolha un Prosody Buscar	na área 🗸
Tipos de Exibição	▼ Português
Normal	Nenhum termo encontrado equivalente a "Prosody"
Descritiva	▼ English
Tipos de Consulta Total Tradutor Modular Consultas Externas Corpus NILC Google Answers.com Wikipedia CORTEC	Go back to search results Prosody. Prosodics. n.m/f.s. deals with language expression, comprehension and fluency. Ex.: Prosodic reading, or reading with expression, is considered one of the hallmarks of fluent reading. The major purpose of the study was to learn how reading prosody is related to decoding and Reading comprehension skills Co- hyponyms: comprehension. Corpus: Frequency order position: (4); Term number of occurrencies: (388). Encyclopedic Information: In linguistics, prosody is the rhythm, stress, and intonation of speech. em: prosody - Wikipedia
	Figure 76. Term PROSODY in VoTec's page (just English).

To have an idea how a complete term is viewed in the system, let's take a look:

Vocabulário Técnico Online	Tela Cheia English Ajuda
Linguística V Escolha uma área V	
Buscar	
Tipos de Exibição	V Português
Normal	<u>/oltar ao resultado da busca</u>
Descritiva	Gramática. Sintaxe. s.f.s. conjunto de regras internalizadas na mente dos indivíduos de uma determinada
Tipos de Consulta Total Tradutor Modular	comunidade linguística; uma teoria sobre uma língua particular; deve refletir a maneira como o falante constrói enunciados. Ex.: a gramática de uma língua natural é uma teoria sobre a Língua-I de um indivíduo <i>dipônimo de:</i> língua. <i>Co-hipônimos:</i> regras internalizadas; teoria sobre a língua. Córpus: <i>Posição na Ordem</i> <i>de Freqüência:</i> (93); N ^o <i>de Ocorrências do termo:</i> (306). Informações Enciclopédicas: Gramática é o conjunto de regras indivíduais usadas para um determinado uso de uma língua, não somente da norma culta, mas também de variantes não padrão. É ramo da Linguística que tem por objetivo estudar a forma, a
Consultas Externas Corpus NILC	composição e todas as questões adicion Em: Gramática - <u>Wikipedia</u>
Google	▼ English
Answers.com Wikipedia	Go back to search results
Wikipedia CORTEC	Grammar . Syntax. Grammar. n.m/f.s. inventory which aims at the building of the language structure; allows mappings between meanings and signals incorporating meaning into the utterance; it is an internally structured set of rules, autonomous of meaning, shaped and reshaped through language use, responding to communicative challenges. Ex.: We describe the grammar is structured internally, and how it adds structure to utterances and decodes it again. <i>Synonyms</i> : principles. <i>Hyponym of</i> : meaning; uterrances; language; system; communication. <i>Hypernym of</i> : mappings; signals; constructions; form-meaning pairings. Corpus : <i>Frequency order position</i> : (3); <i>Term number of occurrencies</i> : (2278). Encyclopedic Information : In linguistics, grammar is the set of structural rules that govern the composition of clauses, phrases, and words in any given natural language. The term refers also to the study of such rules, and this field includes morphology, syntax, and phonology, of em: <i>Grammar</i> - <u>Wikipedia</u>
25/05/2014 08:30 © 2007 Guilherme Fromm - ICMC Jr. Termo elaborado por Virgínia do Nascimento Peixoto (pt) Marcio Issamu Yamamoto (en)	

Figure 77. Term GRAMMAR in the VoTec.

VoTec has many ways to display the information from the data bank. Try them!

REFERENCES



ANTHONY, L. AntConc version 3.4.1. Disponível em: http://www.antlab.sci.waseda. ac.jp/index.html. Acessado em: 20 abril 2014.

AUBERT, F. H. Introdução à metodologia da pesquisa terminológica bilíngüe. São Paulo: Humanitas, 1996.

BERBER SARDINHA, T. Linguística de Corpus. Barueri: Manole, 2004.

FROMM, G. VoTec: a construção de vocabulários eletrônicos para aprendizes de tradução. São Paulo, 2007. Tese (Doutorado – Programa de Pós-Graduação em Estudos Linguísticos e Literários em Inglês – Departamento de Letras Modernas). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

FROMM, G; YAMAMOTO, M. I. Terminologia, terminografia, tradução e linguística de *corpus*: a criação de um vocabulário bilíngue sobre linguística. TAGNIN, S. E. O.; BEVILACQUA, C. (org.). **Corpora na terminologia**. São Paulo: Hub Editorial, 2013.

MCGLASHAN,M. ESRC Centre for CorpusApproaches to Social Science (CASS). **Corpus**: some key terms.Lancaster: Lancaster University, 2013. Disponível em: http:// cass.lancs.ac.uk/?page_id=956. Acesso em: 26 abril 2014.

SCOTT, M. WordSmith Tools version 6. Liverpool: Lexical Analysis Software, 2014.

SUGGESTED BOOKS

BIBER, D; CONRAD, S; REPPEN, R. **Corpus linguistics**: investigating language structure and use. Cambridge: Cambridge University Press, 2002.

BIDERMAN, M. T. C. (1978) **Teoria Lingüística**: teoria lexical e linguística computacional. 2. ed. São Paulo: Martins Fontes, 2001. 356 p. (Coleção Leitura e Crítica).

DELGADO, H. K., FINATTO, M. J., PERNA, C. L. (org). Linguagens especializadas em corpora: modos de dizer e interfaces de pesquisa [recurso eletrônico]. Porto Alegre: EDIPUCRS, 2010. Disponível em: http://www.pucrs.br/edipucrs/ linguagensespecializadasemcorpora.pdf

MCENERY, T. **Corpus linguistics**: an introduction. Edinburgh: Edinburgh University Press, 2001.

SARDINHA, T. B. (org). A língua portuguesa no computador. Campinas: Mercado de Letras, 2005.

SARDINHA, T. B. **Pesquisa em linguística de Corpus com WordSmith Tools**. Campinas: Mercado de Letras, 2009.

SINCLAIR, J. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991.

VIANA, V.; TAGNIN, S. E. O. (Orgs.).**Corpora no ensino de língua estrangeira**. São Paulo: Hub Editorial. 2010.

TAGNIN, S. E. O. (org). Cadernos de Tradução/UFSC/NUT. nº 9. Florianópolis: NUT, 1996. Disponível em: http://www.periodicos.ufsc.br/index.php/traducao/issue/view/432