



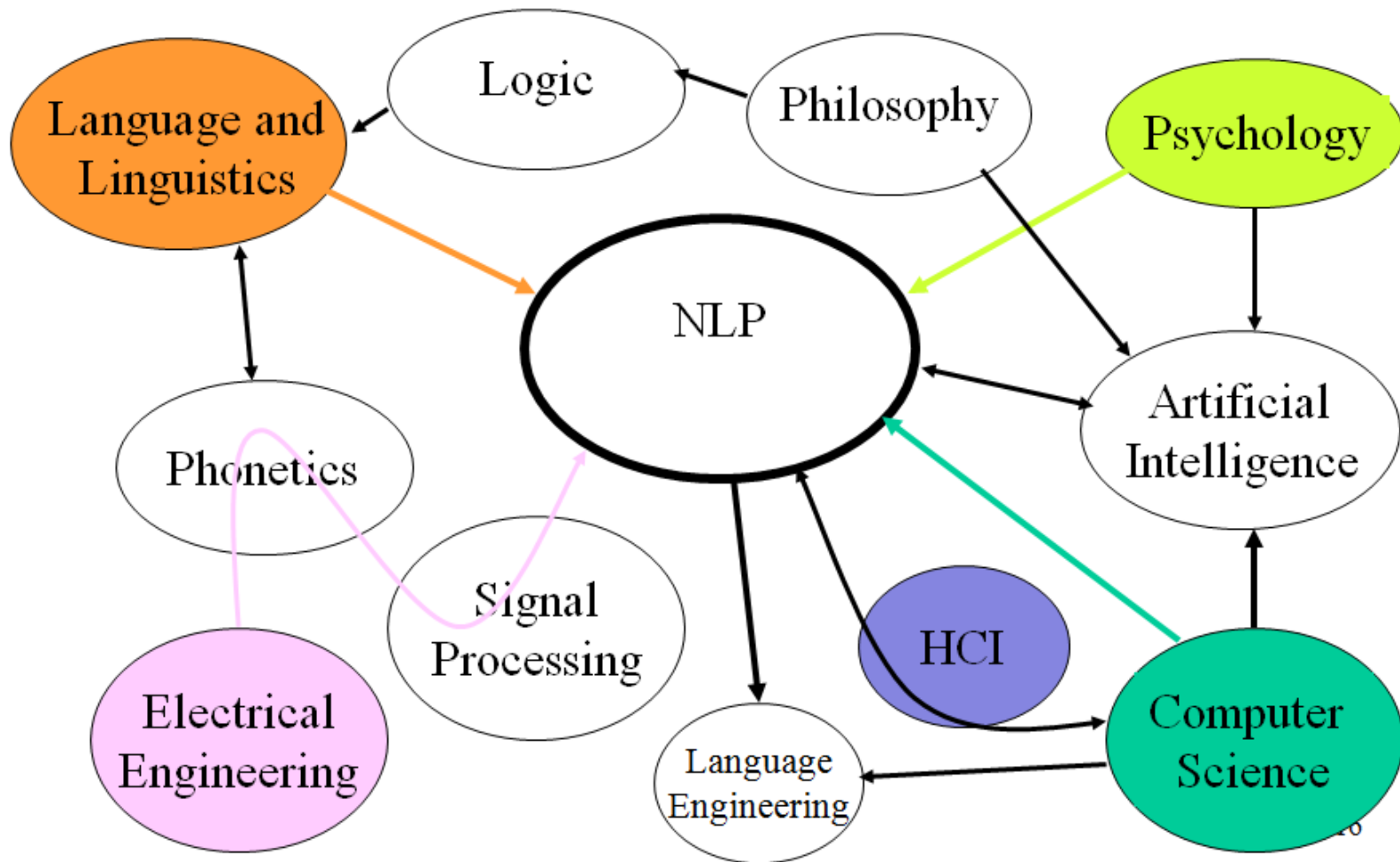
Click to edit Master title style

Computational Linguistic.
Introduction

Mihaela Colhon, 2017



NLP (Computational Linguistic) & HCI



Mihaela Colhon, 2017



Motto ...

Human language is the most exciting and demanding puzzle.



Mihaela Colhon, 2017



The Scope

- Teach computers to communicate with people.

Because ...

- Friendly programs should listen and speak.



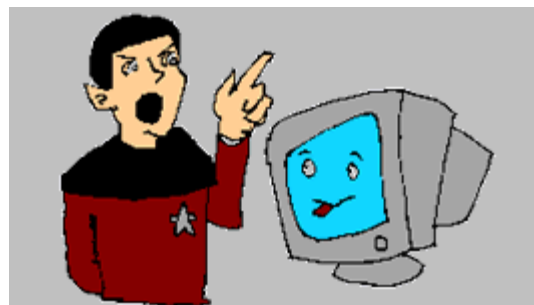
Mihaela Colhon, 2017

Oamenii și-au dorit dintotdeauna să construiască mașini „inteligente” care să le ușureze viața și să comunice cu ele eventual prin limbaj natural.

În prezent calculatoarele (și roboții) sunt considerate cele mai „inteligente” mașini, dar ele sunt departe de a fi atât de inteligente precum și-au dorit oamenii.

În încercarea de a crea inteligență artificială informaticienii au „programat” calculatoarele să imite oamenii: le-au „dotat” cu cunoștințe dintr-un domeniu anume și cu un algoritm pentru o funcție asociată, sperând ca rezultatele calculatorului să egaleze sau să depășească rezultatele oamenilor în acel domeniu.

O tendință actuală în informatică este de a utiliza calculatorul pentru a permite realizarea dialogului dintre om și calculator în maniera cea mai naturală - limbajul. De această problemă se ocupă *Lingvistica computațională*, ramură a *Inteligenței artificiale*.





Scop

Sistemele NLP au fost realizate pentru a extrage informații din baze de date, pentru a controla roboții, pentru a interacționa într-un mod cât mai natural cu utilizatorii, mai precis, pentru a interacționa într-un mod cât mai “prietenos” cu utilizatorul în vederea rezolvării anumitor sarcini sau probleme.

Limbajul natural asigură cea mai “imediată” metodă de interacțiune om-mașină, prezentând mai multe avantaje, unele dintre ele bazându-se pe vocabularul disponibil imediat, care nu implică o pregătire prealabilă.



The Scope

- Language is the fabric of the web.



Mihaela Colhon, 2017



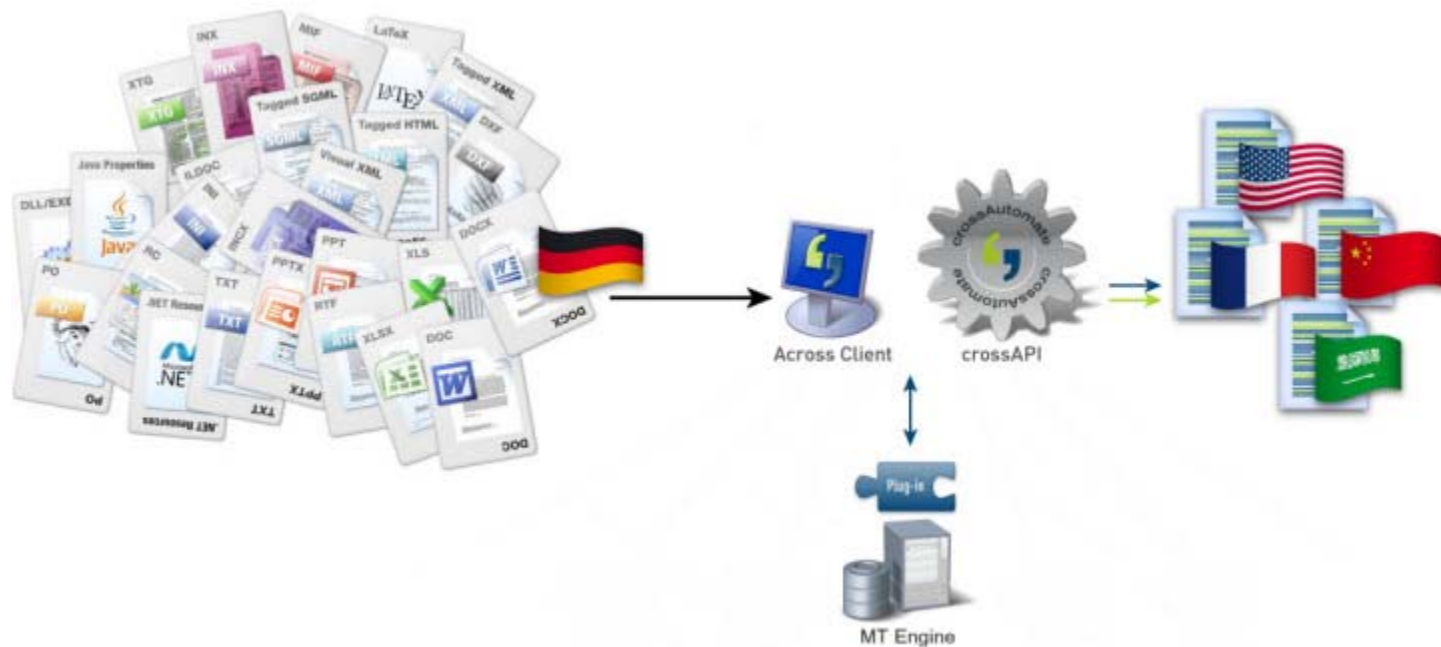
Seven Core Areas of CL

- Machine Translation
- Speech Recognition
- Text-to-Speech
- Natural Language Generation
- Human-Computer Dialogs
- Information Retrieval
- Computational Modeling



Machine Translation (MT)

To use computers to automate some or all of translating from one language to another.





Speech Recognition (SR)

Taking spoken language as input and outputting the corresponding text.





Text-to-Speech (TTS)

Taking text as input and outputting the corresponding spoken language.

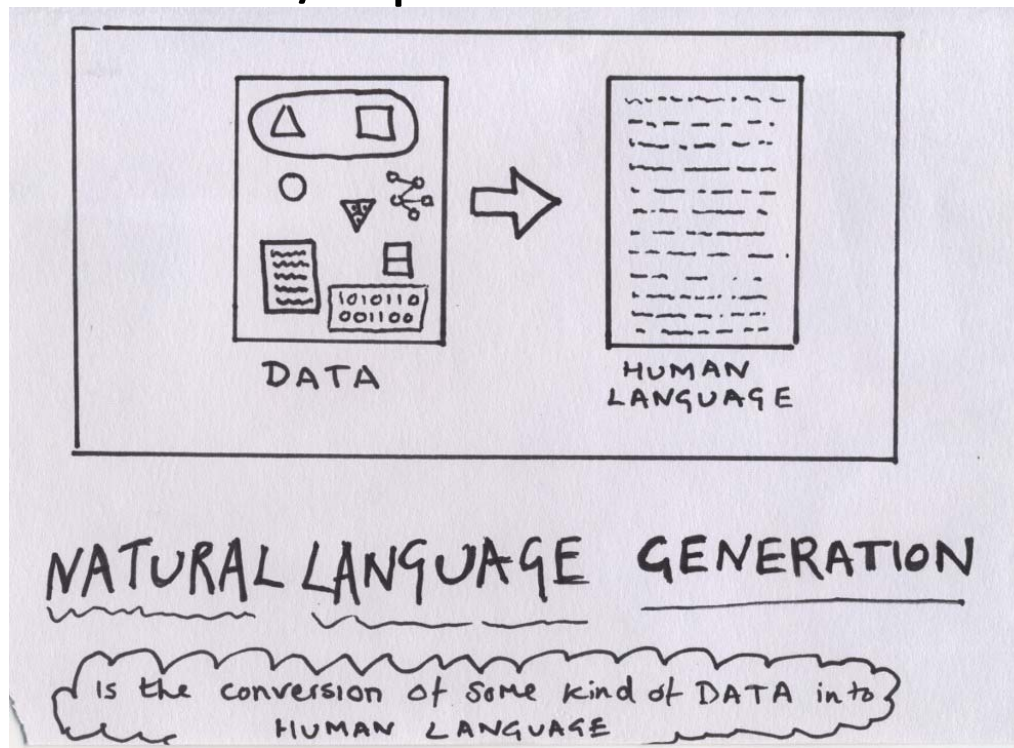
tts process





Natural Language Generation (NLG)

Constructing linguistic outputs from non-linguistic inputs; the NLG goal is to produce natural language from internal data/representation.



Mihaela Colhon, 2017



Human-Computer Dialogs (I)

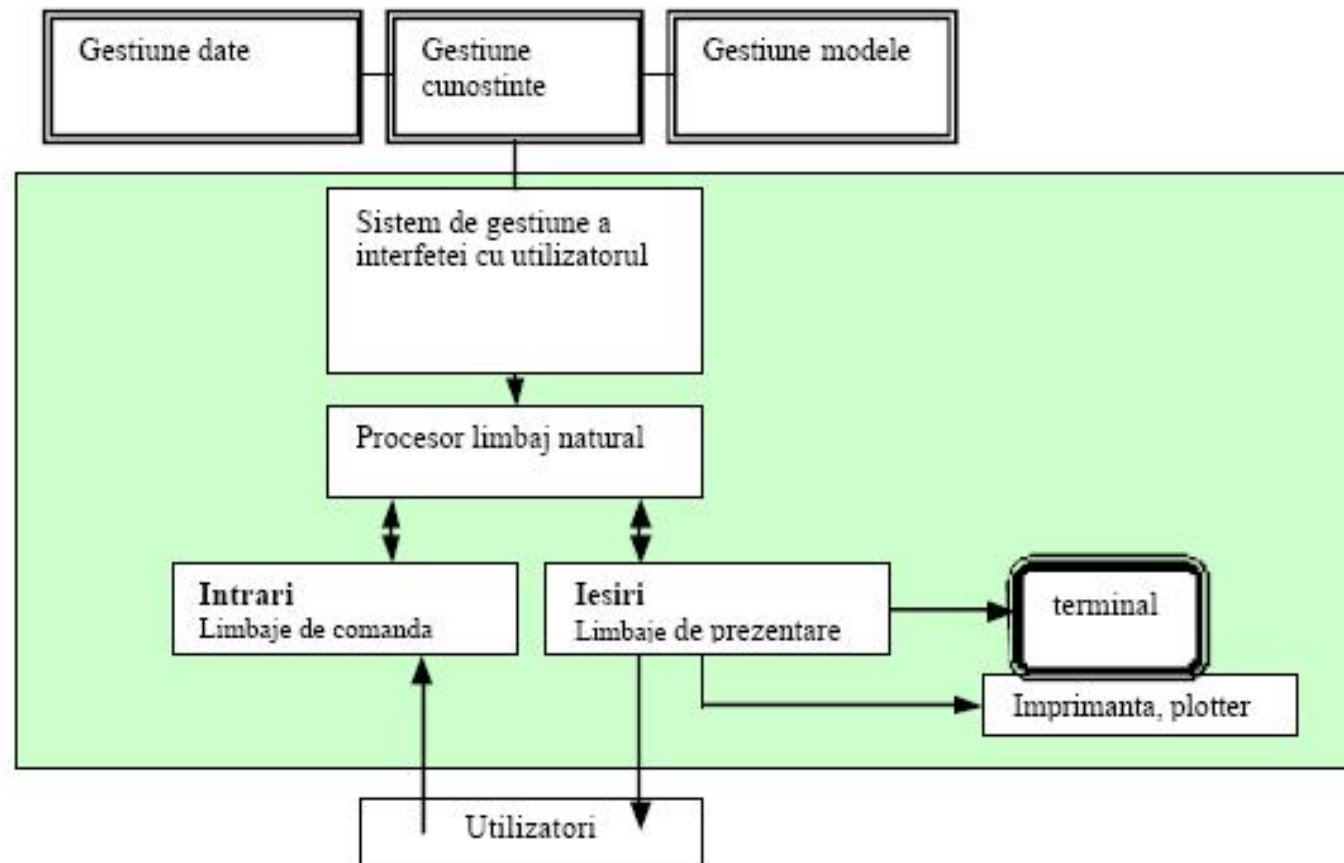
Uses a mix of SR, TTS, and pre-recorded prompts to achieve some goal.

Sistemele de dialog inteligent sunt programe care accepta intrari in limbaj natural si produc in consecinta raspunsuri in limbaj natural. Pentru acesta, urmatoarele module sunt necesare:

- Natural Language Understanding (reprezentarea sensului propozitiilor intr-un anumit limbaj de reprezentare)
- Natural Language Generation
- Dialogue Management



Human-Computer Dialogs (II)





Dialogue System. Case Study

- *What would you like to order?*
a large pizza please
- *What toppings would you like?*
cheese
- *What else would you like?*
a large beer
- *What else would you like?*
nothing else thank you
- *Thank you for ordering one large pizza with cheese and one large beer and nothing else. Your order will be ready in two minutes.*

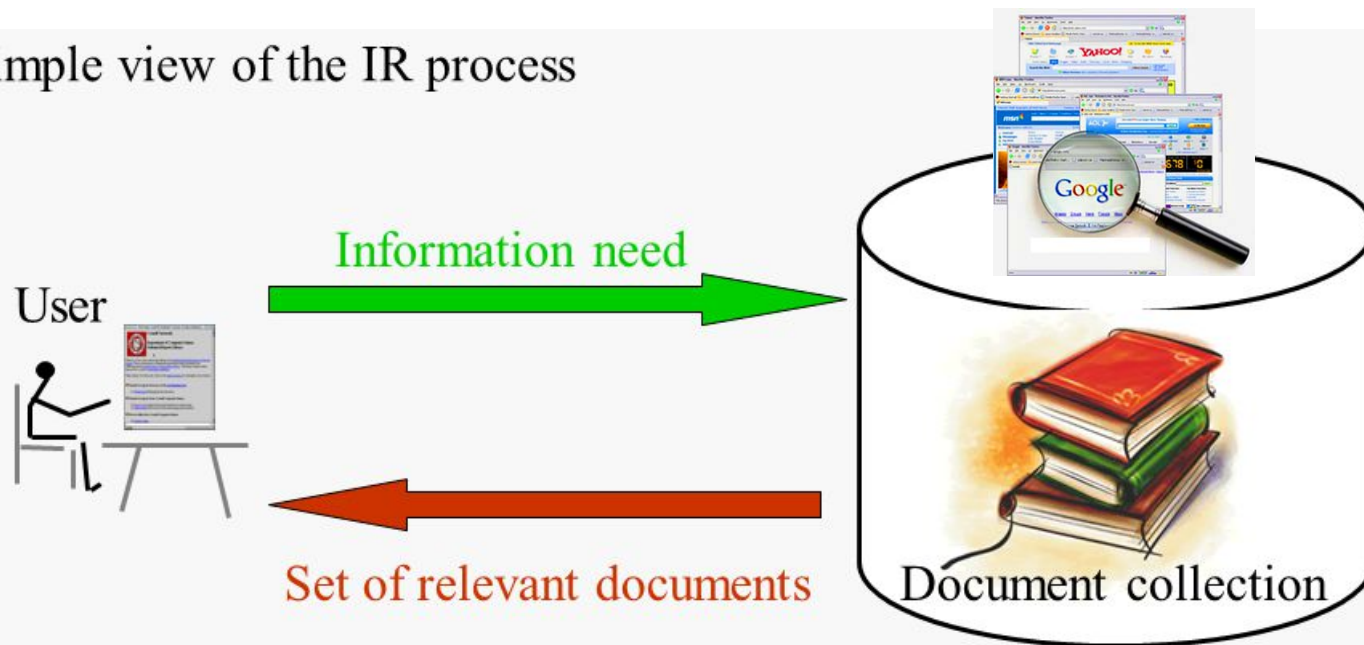




Information Retrieval

Storage, analysis, and retrieval of text documents.

Simple view of the IR process



The set of documents in the answer MUST be relevant to the user's information need. Otherwise the IR process results in complete failure.

Mihaela Colhon, 2017



Computational Modeling

Computational approaches to problem solving, modeling, and development of theories.



Mihaela Colhon, 2017



Prelucrarea limbajului natural

Prelucrarea limbajului natural cuprinde atât înțelegerea cât și generarea de mesaje în limbaj natural. Deoarece înțelegerea implică identificarea structurii sintactice a mesajelor, obținem că prelucrarea limbajului natural nu se poate face fără mecanisme puternice de reprezentare și procesare.





Cum computerele nu posedă mentalitatea umană, prin definiție nu pot înțelege. Bineînțeles, Procesarea Limbajului Natural (Natural Language Processing – NLP) este o problemă generală și pentru a fi mai specifici, o vom separa în categorii conform nivelului crescut sau complexității acestei procesări:

- Morfologie și procesare morfologică
- Sintaxă și procesarea sintactică
- Semantică și procesarea semantica



Nivele de analiză în NLP (I)

- **Fonetic:** formarea cuvintelor din sunete
- **Morfologic:** formarea cuvintelor din morfeme (cea mai mică unitate din structura morfologică a cuvântului cu un sens determinat (lexical sau gramatical))
- **Sintactic:** combinarea cuvintelor în fraze și funcțiile pe care acestea le iau în cadrul combinațiilor
- **Semantic:** înțelesul frazelor din înțelesul cuvintelor
- **Pragmatic:** înțelesul frazelor în funcție de contextul în care apar



Nivele de analiză în NLP (II)





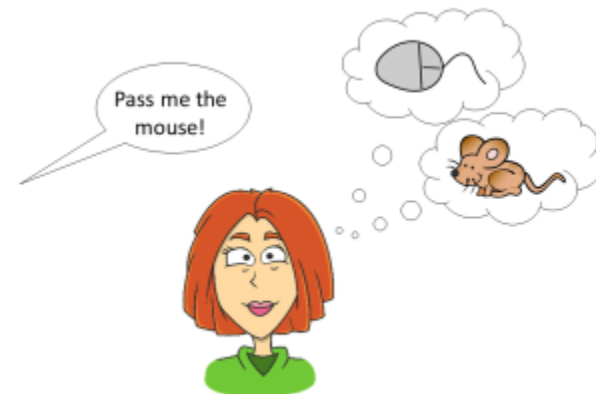
Problem with language (in general)

- Words are ambiguous (*bank, round, take*)
- Sentences are ambiguous

The chicken is ready to eat

The man saw the girl with a telescope

Remove bulb, cover, and replace





Pragmatic problems

- We don't always say what we mean

Can you pass the salt?

I'm sorry (= Say it again)

Do you want some more? You're alright.

- We don't always mean what we say

It's raining cats and dogs



Mihaela Colhon, 2017

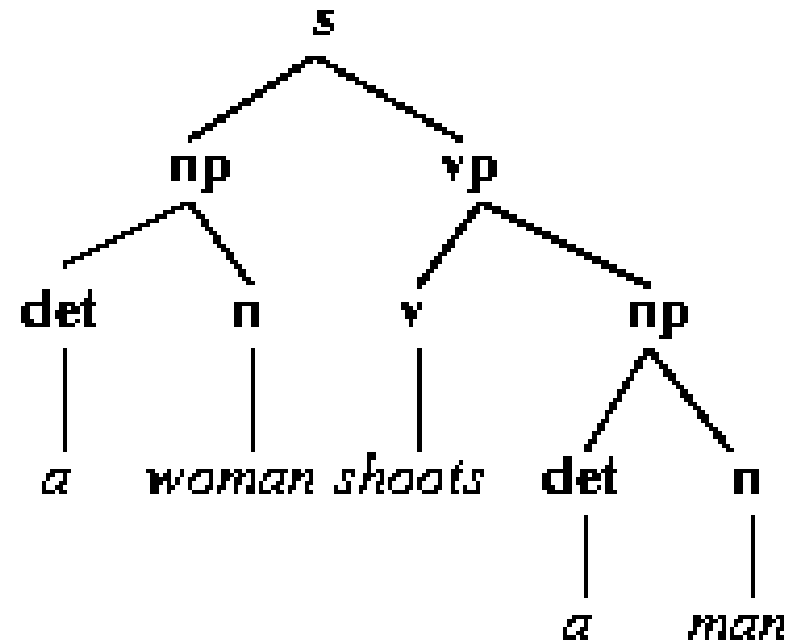


What is syntax? Context-Free Grammars

O gramatica context-free este un sistem:

$$G=(V_N, V_T, S, P)$$

- V_N este mulțimea simbolurilor neterminale
- V_T este mulțimea simbolurilor terminale
- $S \in V_N$ este simbolul inițial al gramaticii
- $P \subset V_N \times (V_N \cup V_T)^*$ este mulțimea producțiilor





Corpus. Utilizări

Probabil, cea mai importantă tendință în lingvistică este utilizarea masivă de corpusuri mari de texte în limbaj natural. În orice sistem NLP, corpusurile sunt adeseori utilizate pentru a furniza date empirice și statistice.

În mod obișnuit, aplicațiile de NLP care utilizează corpusurile ca o resursă lingvistică de bază sunt programele pentru dezambiguizarea sensurilor (în engleza Word Sense Disambiguation - WSD) și toate tipurile de parsere.



Corpus

În lingvistica computațională, un corpus este o colecție structurată de texte, stocate electronic și reprezentate într-un anumit formalism pentru a putea fi ușor “citite” de aplicațiile specializate în extragerea de informații din aceste colecții.

Un corpus poate conține texte într-o singură limbă - *corpus monolingv*, sau texte în mai multe limbi - *corpus multilingv*. Corpusurile multilingve în care textele sunt grupate în baza existenței unor echivalenți de traducere se numesc *corpusuri paralele*.





Corpus

Corpusurile reprezintă **baza de cunoștințe** a oricărei aplicații din domeniul NLP. Analizarea și procesarea cât mai multor tipuri de corpusuri este subiectul multor lucrări din *Lingvistică Computațională*, *Recunoașterea Vorbirii* (unde corpusurile sunt adesea folosite pentru a crea modele Markov ascunse).

În scopul de a face corpusurile utile pentru cercetarea lingvistică, acestea sunt adeseori supuse unui proces numit **adnotare**. Un exemplu de adnotare a unui corpus este identificarea părții de vorbire pentru fiecare cuvânt din corpus, unde informația despre partea de vorbire (verb, substantiv, adjectiv, etc.) este atașată fiecărui cuvânt sub forma unei etichete corespunzătoare.



Analiză sintactică. Parsing

Un **program de parsing** este o căutare în cadrul unui enunț a unor posibile caracteristici structurale ale acestuia, caracteristicile fiind constrânse de faptul că ele trebuie să fie compatibile cu o anumită secvență de cuvinte.

Parsingul sintactic încearcă convertirea expresiilor din limbaj natural în structuri de arbori sau legături de dependență ce prezintă structura sintactică a expresiei. Aceste structuri sintactice sunt expediate unui interpretator semantic pentru o procesare ulterioară.



Adnotarea morfo-sintactică a textelor (I)

Adnotarea, sau etichetarea morfosintactică (în engleza, Part Of Speech tagging – POS tagging), este procesul de alăturare a informației ce decurge din analiza morfologică și sintactică a unui corpus de texte. Una din sarcini → atribuirea pentru fiecare cuvânt din corpus a unei etichete de parte de vorbire (POS tag).

Procesul de adnotare morfo-sintactică trebuie să acorde atenție contextului în care apare cuvântul care se adnotează, deoarece categoria gramaticală a unui același cuvânt poate depinde de construcția în care acesta apare: *gust* interesant (substantiv), eu *gust* (verb, persoana I sg).

Un tagger trebuie să delimiteze un cuvânt, să determine posibilele lui interpretări și să o atribuie pe cea corectă, conform contextului. Adică să rezolve ambiguitatea.



Adnotarea morfo-sintactică a textelor (II)

Parte de vorbire	Atribute morfologice
Verb	Timp, persoana, numar, gen
Substantiv	Numar, gen, tip (comun sau propriu)
Adjectiv	Numar, gen, grad (pozitiv, posesiv, interogativ, demonstrativ, nedefinit, relativ)
Pronume	Tip (personal, posesiv, interogativ, demonstrativ, nedefinit, relativ), gen, numar, caz
Determinant	Numar, gen, tip (articulat, posesiv, demonstrativ, interogativ)
Prepozitie	Tip (loc, timp, mod)
Adverb	Tip (loc, timp, modalitate)
Numeral	Tip (cardinal, ordinal), numar
Conjunctie	Tip (coordonatoare sau subordonatoare)

Mihaela Colhon, 2017



i Adnotarea morfo-sintactică a textelor. (III) Proiectul MULTEXT-EAST

În cadrul proiectului european MULTEXT-EAST au fost create coduri cu descrierile detaliate pentru POS tag-uri aplicând formalismul atribut / valoare.

Pentru limba engleza s-au creat 83 de etichete, iar pentru limba româna inițial s-au definit 614 taguri, dar în final numărul de taguri a fost redus la 262.



i
**Adnotarea morfo-sintactică
 a textelor. (IV)**
Proiectul MULTEXT-EAST

**TAGURI MULTEXT-EAST.
 N - SUBSTANTIV (NOUN)**

No.	Field	Value	EN	RO
1.	Type	common (c)	x	x
		proper (p)	x	x
2.	Gender	masculine (m)	x	x
		feminine (f)	x	x
		neuter (n)	x	x
3.	Number	singular (s)	x	x
		plural (p)	x	x
4.	Case	vocative (v)		x
		direct (r)		x
		oblique (o)		x
5.	Definiteness	no (n)		x
		yes (y)		x

Exemple:

- frate **Ncmsrn**
- fratelui **Ncmsoy**
- surorile **Ncfpry**
- București **Npmsrn**

**TAGURI MULTEXT-EAST.
 V - VERB (VERB)**

No.	Field	Value	EN	RO
1.	Type	main (m)	x	x
		aux (a)	x	x
		modal (o)	x	x
		copula (c)		x
		base (b)	x	
2.	VForm	indicative (i)	x	x
		subjunctive (s)		x
		imperative (m)		x
		conditional (c)	x	
		infinitive (i)	x	x
		participle (p)	x	x
		gerund (g)		x
3.	Tense	present (p)	x	x
		imperfect (i)		x
		future (f)		
		past (s)	x	x
		pluperfect (l)		x
4.	Person	first (1)	x	x
		second (2)	x	x
		third (3)	x	x
5.	Number	singular (s)	x	x
		plural (p)	x	x
6.	Gender	masculine (m)	x	x
		feminine (f)	x	x
		neuter (n)	x	x

Exemple:

- abandonam **Vmii1s**
- abandună **Vmis3s**
- trebuie **Voip**
- sunt **Vcip1s**

Mihaela Colhon, 2017



Exemplu de marcatori pentru informații morfo-sintactice

id ₁		avand
lema avea	ana 1+,vmg	
id ₂		in
lema in	ana 5+,sp sa	
id ₃		vedere
lema vedere	ana 1+,ncfsm	
id ₄		art.
lema art.	ana 8+,yn	
id ₅		54
lema 54	ana 1+,mc	
id ₆		din
lema din	ana 5+,sp sa	
id ₇		tratat
lema tratat	ana 1+,ncms-n	

© 2015 Mihaela Colhon



What is semantics?

The study of the linguistic meaning of morphemes, words, phrases, and sentences.

One way to represent semantic properties is by means of semantic features.

Woman	Father	Girl
+female	+male	+female
+human	+human	+human
-young	+parent	+young



Study Case AntConc

http://www.laurenceanthony.net/antconc_index.html



Laurence Anthony's Website

Home Resume Publications Software Classes Photo Albums Links Contact

AntConc Homepage

Latest Release



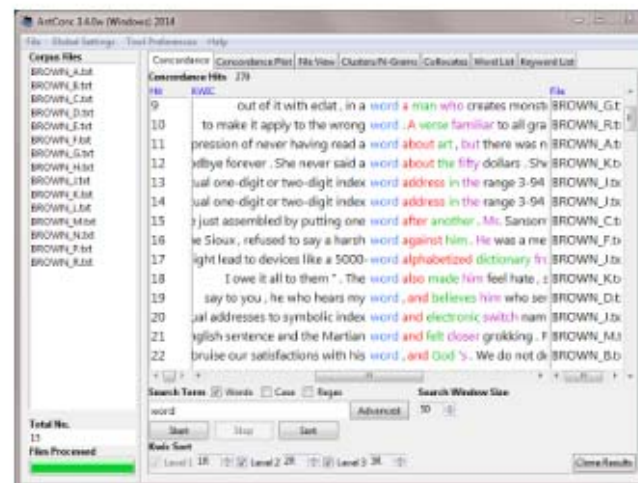
AntConc

A freeware corpus analysis toolkit for concordancing and text analysis.

[AntConc Homepage] [Screenshots] [Help]

Downloads:

- Windows (3.4.4)
- Macintosh OS X 10.7-10.12 (3.4.4)
- Macintosh OS X 10.6 (3.4.1)
- Linux (3.4.3)
- Older versions





FileView

AntConc 3.4.3w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files
Senseval3.RomanianLS

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

File View Hits 0 File Senseval3.RomanianLS.txt

In ce priveste zona de provenienta , s - a constatat ca principalul furnizor este Moldova , aproximativ 30 % dintre ei provin de aici si numai 14,5 % din Bucuresti .
Pentru ca din 90 incoace , prea putina lume s - a ocupat de soarta lor , copiii strazii nu mai sint copii .
Un sfert dintre cei care dorm in canale au virste mai mari de 16 - 18 ani .
Marea masa a copiilor strazii o reprezinta baietii , peste 70 % , fetele fiind reprezentative in principal in grupa de virsta 16 - 17 ani .
Saracia , certurile si violenta in familie , dar si dezmembrarea familiei sint principalele cauze pentru care copilul fuge de acasa .

El a spus ca pozele vor fi trimise la Politia Rutiera care ii va amenda pe soferi .
Constantin Sofroni a declarat ca fotografiile vor fi realizate de sefii echipelor de maturatori care vor fi dotati cu aparate de fotografiat achizitionate de firmele care se ocupa de salubritatea orasului .
Originalii fotografi vor trebui sa depuna zilnic la primarie situatia masinilor parcate ilegal , precum si situatia cosurilor de gunoi si a capacelor de canal furate .
Constantin Sofroni s - a plins de faptul ca " 30 % din suprafata asfaltata a Sucevei

Search Term Words Case Regex Hit Location

Advanced 0

Start Stop

Clone Results

Total No.
1
Files Processed



Concordance

AntConc 3.4.3w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files
Senseval3.RomanianS

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 44

Hit	KWIC	File
1	urilor de gunoi si a capacelor de canal furate . Constantin Sofron	Senseval3.Ro
2	filme ". Capacul de la gura de canal , cit o anvelopa de tractor ,	Senseval3.Ro
3	postului . Gazul s - a infiltrat in canal si , de - aici , in adapost . E	Senseval3.Ro
4	citorii din serviciile locale - apa , canal , salubritate - si nici cei de	Senseval3.Ro
5	e au fost dirijate printr - un nou canal spre Bahlui , fiind evacuate	Senseval3.Ro
6	nite in continuare , macar pe un canal . Simbata noaptea , aviane	Senseval3.Ro
7	aminte purtate anterior intr - un canal de scurgere , plin cu apa .	Senseval3.Ro
8	aze au fost inchise Regia de Apa Canal Craiova are si ea de recup	Senseval3.Ro
9	gitul si s - a aruncat intr - un canal Un sucevean si - a batut fa	Senseval3.Ro
10	gitul si s - a aruncat intr - un canal TIBERIU AVRAM Duminica	Senseval3.Ro
11	gitul si s - a aruncat intr - un canal In cursul zilei de ieri , in ju	Senseval3.Ro
12	i , in jurul orei 10.30 , dintr - un canal aflat in apropierea Fabricii	Senseval3.Ro
13	scutit dupa care s - a aruncat in canal , murind inecat . Organele	Senseval3.Ro
14	gitul si s - a aruncat intr - un canal Un sucevean si - a batut fa	Senseval3.Ro

Search Term Words Case Regex Search Window Size 50

canal Advanced

Start Stop Sort

Kwic Sort

Level 1 1R Level 2 2R Level 3 3R

Clone Results

Total No. 1
Files Processed



Left Concordance

AntConc 3.4.3w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files
Senseval3.RomanianL9

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 44

Hit	KWIC	File
1	de lei Regia Judeteana de Apa - Canal Iasi a anuntat noile pretur	Senseval3.Ro
2	aze au fost inchise Regia de Apa Canal Craiova are si ea de recup	Senseval3.Ro
3	Regia Autonoma Judeteana Apa - Canal s - a aflat in postura penib	Senseval3.Ro
4	ditorii din serviciile locale - apa , canal , salubritate - si nici cei de	Senseval3.Ro
5	urilor de gunoi si a capacelor de canal furate . Constantin Sofron	Senseval3.Ro
6	in filme " . Capacul de la gura de canal , cit o anvelopa de tractor ,	Senseval3.Ro
7	ce a intrat intr - o gura de canal Nemteanul Nemteanu si -	Senseval3.Ro
8	ce a intrat intr - o gura de canal DANIEL CONTESCU Marti	Senseval3.Ro
9	ce a intrat intr - o gura de canal Fotbalistul Mihai Nemtean	Senseval3.Ro
10	mt , incercind sa evite gurile de canal , cind , la un moment dat ,	Senseval3.Ro
11	la RGAB ca sa ridice gurile de canal la nivelul strazii . Muncito	Senseval3.Ro
12	ea problema a sudarii gurilor de canal , lucru pe care primaria nu	Senseval3.Ro
13	ercat fara succes sa il scoata din canal in cursul zilei de joi . Aces	Senseval3.Ro
14	ercat fara succes sa il scoata din canal in cursul zilei de joi . Aces	Senseval3.Ro

Total No. 1
Files Processed

Search Term Words Case Regex Search Window Size 50

canal [Advanced]

Start Stop Sort

Kwic Sort Level 1 1L Level 2 2L Level 3 3L [Clone Results]



Right Concordance

AntConc 3.4.3w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files
Senseval3.RomanianLS

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 44

Hit	KWIC	File
1	ri distanta de comuna , intr - un canal adiacent riului . Se pare ca	Senseval3.Ro
2	ri distanta de comuna , intr - un canal adiacent riului . Se pare ca	Senseval3.Ro
3	i , in jurul orei 10.30 , dintr - un canal aflat in apropierea Fabricii	Senseval3.Ro
4	i , in jurul orei 10.30 , dintr - un canal aflat in apropierea Fabricii	Senseval3.Ro
5	mt , incercind sa evite gurile de canal , cind , la un moment dat ,	Senseval3.Ro
6	n filme " . Capacul de la gura de canal , cit o anvelopa de tractor ,	Senseval3.Ro
7	etri de Torino , au gasit intr - un canal corpul fara viata al lui Ioan	Senseval3.Ro
8	etri de Torino , au gasit intr - un canal corpul fara viata al lui Ioan	Senseval3.Ro
9	le , citiva sateni au descoperit in canal corpul lui Vasile Cristurear	Senseval3.Ro
10	le , citiva sateni au descoperit in canal corpul lui Vasile Cristurear	Senseval3.Ro
11	aze au fost inchise Regia de Apa Canal Craiova are si ea de recup	Senseval3.Ro
12	ce a intrat intr - o gura de canal DANIEL CONTESCU Marti	Senseval3.Ro
13	minute , in adevarate riuri . Un canal de colectare care n - a fost	Senseval3.Ro
14	vem chiar si localizarea : la est - canal de desecare - 157 m , la ve	Senseval3.Ro

Search Term Words Case Regex Search Window Size 50

canal [Start] [Stop] [Sort] [Advanced]

Kwic Sort Level 1 1R Level 2 2R Level 3 3R [Clone Results]

Total No. 1
Files Processed



Concordance Plot

The screenshot shows the AntConc 3.4.3w (Windows) 2014 application window. The main interface is divided into several sections:

- Corpus Files:** A list on the left showing "Senseval3.RomanianLS" selected.
- Search Term:** "canal" is entered in the search field.
- Options:** "Words" is checked, while "Case" and "Regex" are unchecked.
- Concordance Hits:** 44 hits are displayed.
- Total Plots:** 0 plots are shown.
- HIT FILE: 1 FILE: Senseval3.RomanianLS.txt**
- Plot:** A vertical bar chart representing the concordance plot for the search term "canal".
- Statistics:** "No. of Hits = 44" and "File Length (in chars) = 26980" are displayed.
- Plot Zoom:** Set to "x1".
- Buttons:** "Start" and "Stop" buttons are visible at the bottom.



Clusters

AntConc 3.4.3w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files
Senseval3.RomanianLS

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Total No. of Cluster Types 26 Total No. of Cluster Tokens 44

Rank	Freq	Range	Cluster
1	7	1	canal de
2	4	1	canal corpul
3	4	1	canal in
4	2	1	canal , linga
5	2	1	canal , murind
6	2	1	canal adiacent
7	2	1	canal aflat
8	2	1	canal tiberiu
9	2	1	canal un
10	1	1	canal , cind
11	1	1	canal , cit
12	1	1	canal , lucru
13	1	1	canal , miinile

Search Term Words Case Regex N-Grams Cluster Size
canal Advanced Min. 2 Max. 2

Start Stop Sort Min. Freq. 1 Min. Range 1

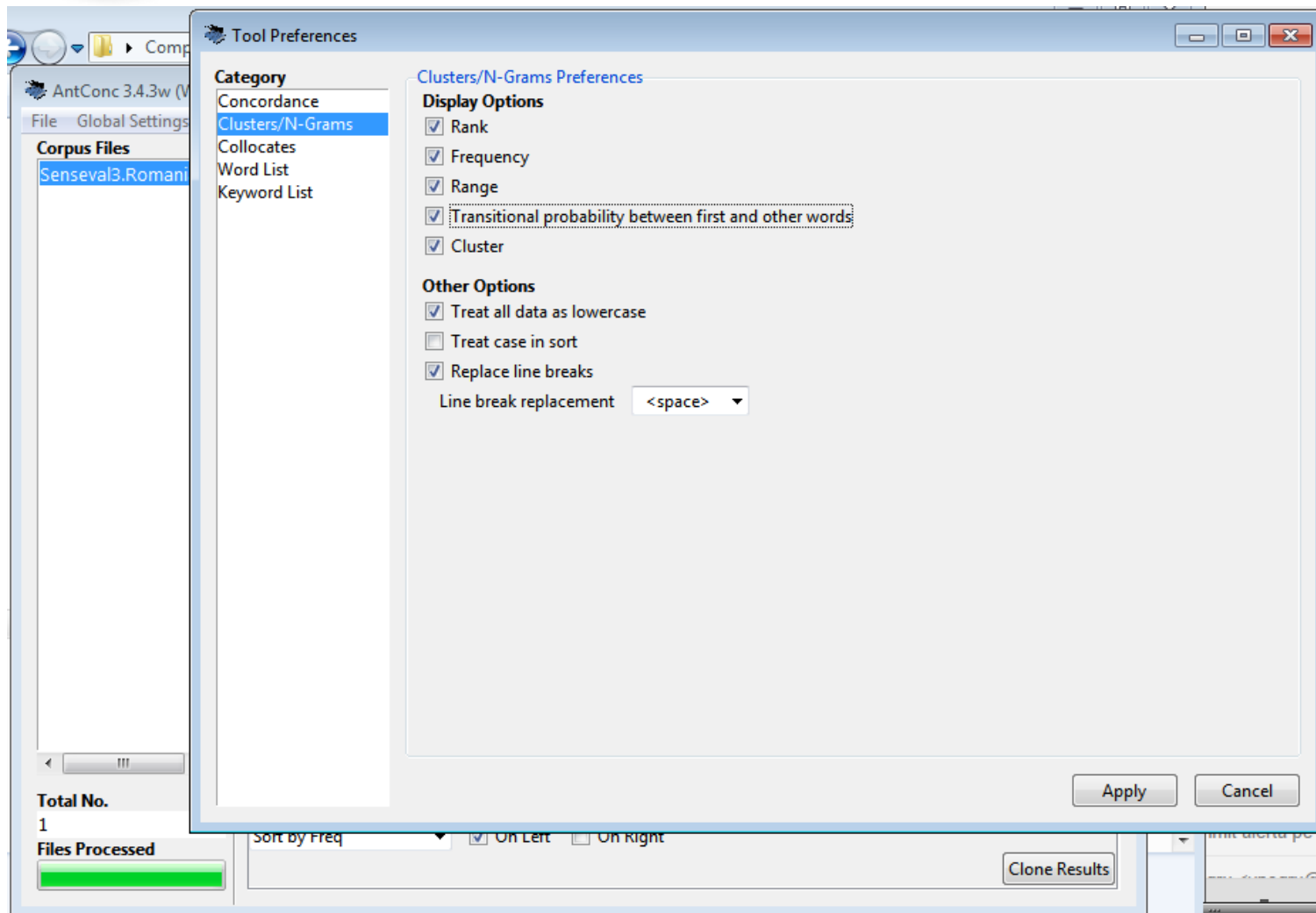
Sort by Invert Order Search Term Position
Sort by Freq On Left On Right

Clone Results

Total No. 1
Files Processed



Clusters Preferences





Clusters

AntConc 3.4.3w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files
Senseval3.RomanianLS

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Total No. of Cluster Types 26 Total No. of Cluster Tokens 44

Rank	Freq	Range	Prob	Cluster
1	7	1	0.159	canal de
2	4	1	0.091	canal corpul
3	4	1	0.091	canal in
4	2	1	0.045	canal , linga
5	2	1	0.045	canal , murind
6	2	1	0.045	canal adiacent
7	2	1	0.045	canal aflat
8	2	1	0.045	canal tiberiu
9	2	1	0.045	canal un
10	1	1	0.023	canal , cind
11	1	1	0.023	canal , cit
12	1	1	0.023	canal , lucru
13	1	1	0.023	canal , miinile

Search Term Words Case Regex N-Grams Cluster Size
canal Advanced Min. 2 Max. 2

Start Stop Sort Min. Freq. 1 Min. Range 1

Sort by Invert Order Search Term Position
Sort by Freq On Left On Right

Total No. 1
Files Processed

Clone Results



Collocates

AntConc 3.4.3w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files
Senseval3.RomanianLS

Concordance Concordance Plot File View Clusters/N-Grams **Collocates** Word List Keyword List

Total No. of Collocate Types: 182 Total No. of Collocate Tokens: 440

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	36	14	22	3.69522	de
2	27	22	5	5.31851	un
3	23	14	9	3.66661	a
4	18	18	0	5.75591	intr
5	18	10	8	3.81340	in
6	13	11	2	5.19896	s
7	9	2	7	2.98773	si
8	8	8	0	6.41606	aruncat
9	7	4	3	2.62516	la
10	6	3	3	3.81340	care
11	6	4	2	3.21675	au
12	6	5	1	4.47051	apa
13	5	4	1	2.73799	din
14	4	0	4	5.12656	zilei

Search Term Words Case Regex

Window Span Same

canal From... 5L To... 5R

Start Stop Sort

Sort by Invert Order

Sort by Freq

Min. Collocate Frequency 1

Clone Results

Total No. 1
Files Processed



Word List

AntConc 3.4.3w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files
Senseval3.RomanianLS

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Word Types: 1487 Word Tokens: 4227 Search Hits: 0

Rank	Freq	Word	Lemma Word Form(s)
1	267	de	
2	174	a	
3	123	in	
4	109	la	
5	109	si	
6	72	din	
7	65	un	
8	62	au	
9	46	fost	
10	45	ca	
11	45	cu	
12	44	canal	
13	41	care	
14	38	ne	

Search Term Words Case Regex

Hit Location

canal 0

Sort by Invert Order

Sort by Freq

Lemma List Loaded

Total No.
1

Files Processed



Keyword List

AntConc 3.4.3w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files

Senseval3.RomanianLS

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List **Keyword List**

Types Before Cut: 1487 Types After Cut: 1487 Search Hits: 0

Rank	Freq	Keyness	Keyword
1	44	0.273	canal
2	62	0.235	au
3	26	0.231	apa
4	34	0.188	dupa
5	20	0.178	ani
6	32	0.171	intr
7	123	0.146	in
8	46	0.117	fost
9	13	0.116	familia
10	12	0.107	ieri
11	11	0.098	vor
12	10	0.089	ale
13	10	0.089	cristurean
14	10	0.089	fiind

Search Term Words Case Regex Hit Location Search Only 0

Start Stop Sort Reference Corpus Loaded

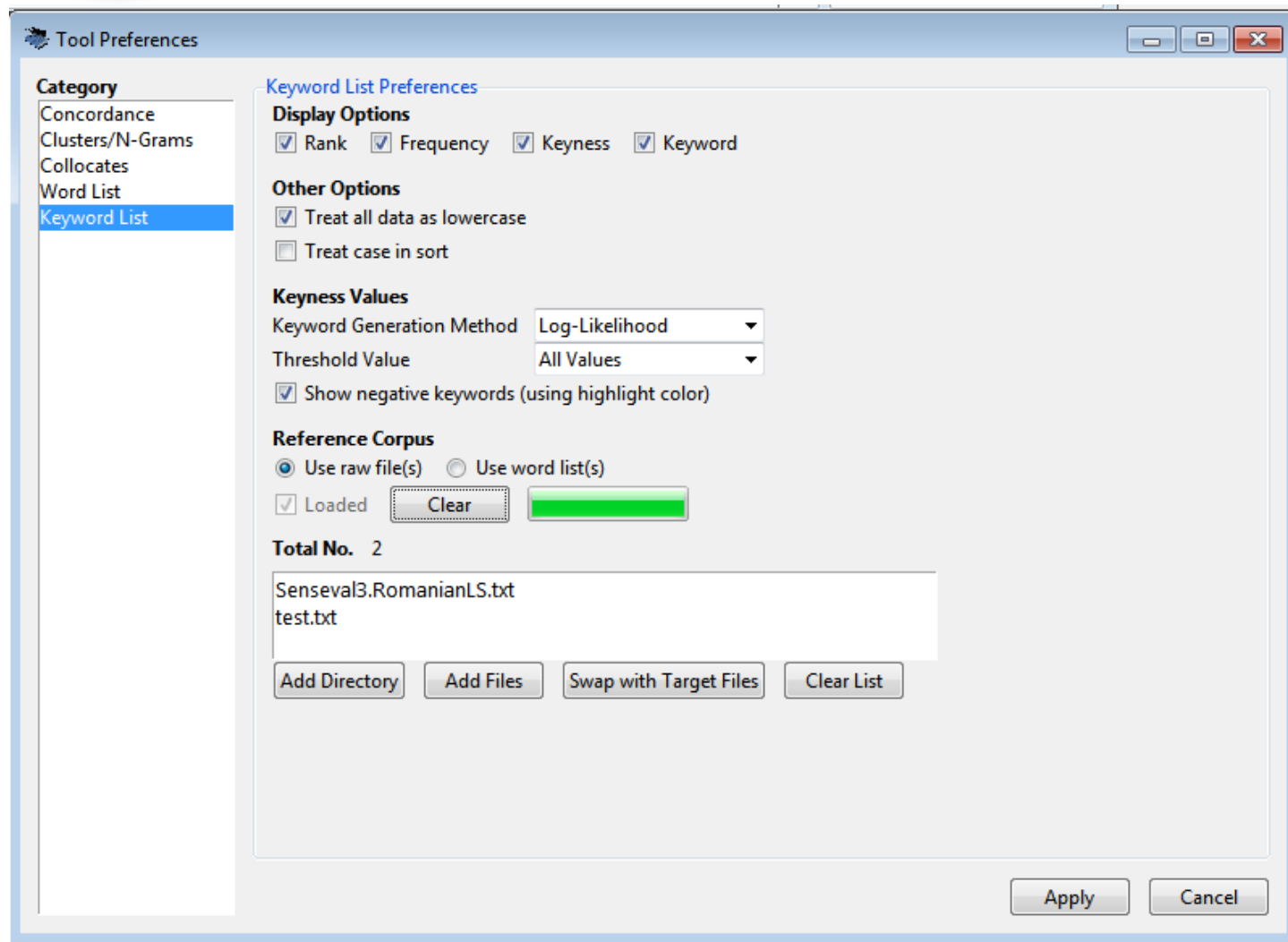
Sort by Invert Order
Sort by Keyness

Total No. 1
Files Processed

Clone Results



Keyword Preferences





Negative Keywords

AntConc 3.4.3w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files

Senseval3.RomanianL...

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List **Keyword List**

Types Before Cut: 1487 Types After Cut: 1487 Search Hits: 0

Rank	Freq	Keyness	Keyword
1378	1	0.000	mea
1379	65	0.000	un
1	1	6.339	e
2	1	1.550	ea
3	1	1.550	fel
4	2	0.959	ajunge
5	2	0.959	cei
6	4	0.879	spre
7	1	0.798	alba
8	1	0.798	calea
9	1	0.798	casa
10	1	0.798	multi
11	1	0.798	prea
12	3	0.659	chiar

Search Term Words Case Regex Hit Location

Advanced Search Only 0

Start Stop Sort Reference Corpus Loaded

Sort by Invert Order

Sort by Keyness

Clone Results

Total No. 1
Files Processed