



# **How to work with text corpora...**

## **... to perform terminology tasks**

Chiara Messina

18-06-2018



# Need analysis

- Get an idea of terminology used in a specific field / by a specific customer
- Term extraction
- Monitoring of changes/shifts in term meaning and use
- Identification of homonyms
- Identification of multi-word terms
- Disambiguation of synonyms
- Entry compilation (Definition / Context / phraseology / collocates)

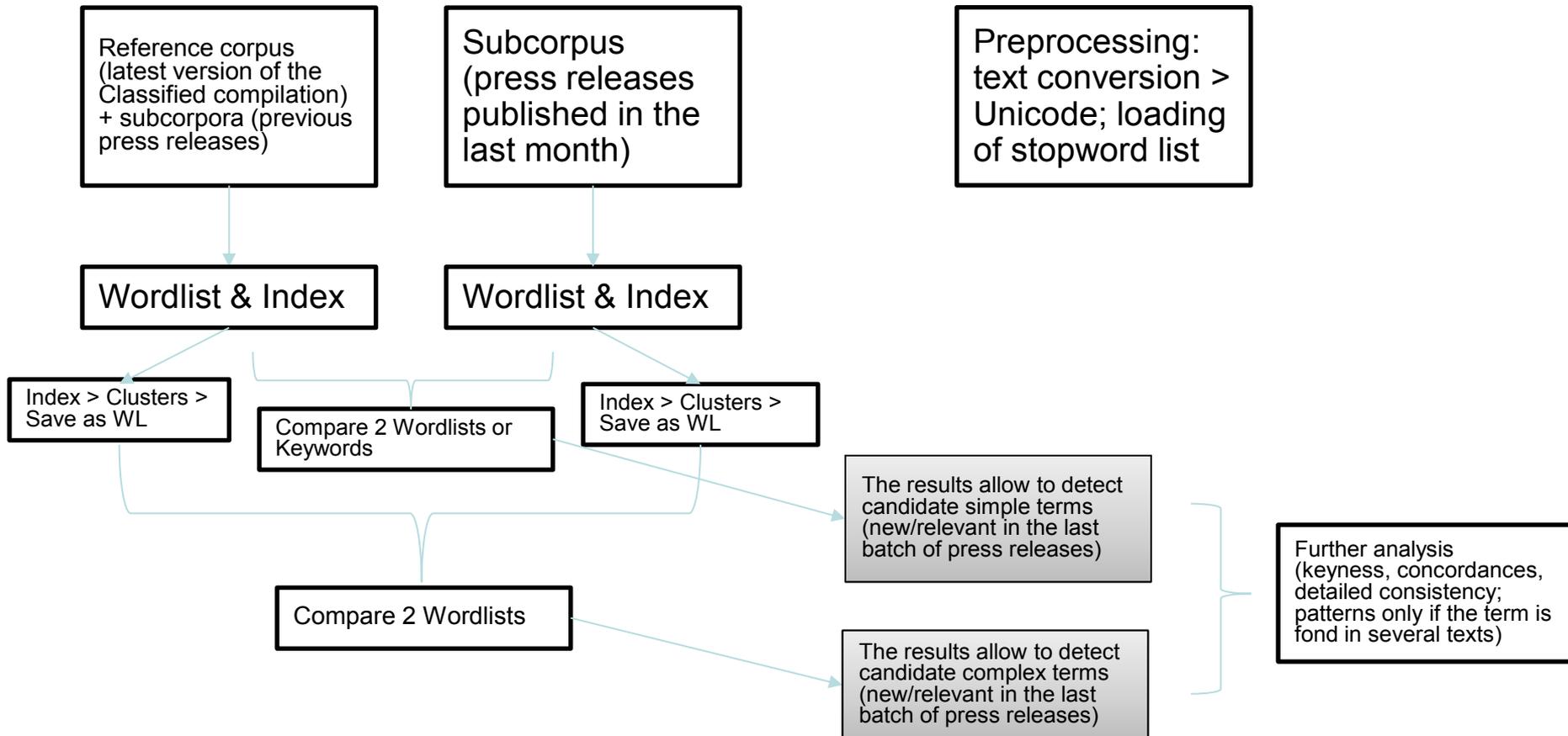


# Tools

Goal	Tool
Get an idea of terminology used in a specific field / by a specific customer	Wordlist, Detailed Consistency
Term extraction	Wordlist / Compare 2 Wordlists
Monitoring of changes/shifts in term meaning and use	Clusters (Index), Collocates, Patterns (Concordance)
Identification of multi-word terms	Clusters (Index), Collocates (Concordance), Wordgrams
Disambiguation of synonyms	Concordance, Detailed Consistency
Entry compilation	According to required data category



# Detect new terms in press releases: workflow





# Corpus building

The screenshot shows a web browser window with the address bar displaying `https://kav-search.bk.admin.ch/terminology/news/`. The browser's address bar also shows the text "Index of /terminology/news". The browser's taskbar at the top contains several open tabs, including "Foraggicoltura-superfici i...", "Raccolta sistematica RS in...", "fscvext", "TERMDAT", "GEVER BK", and "Procedure di consultazion...".

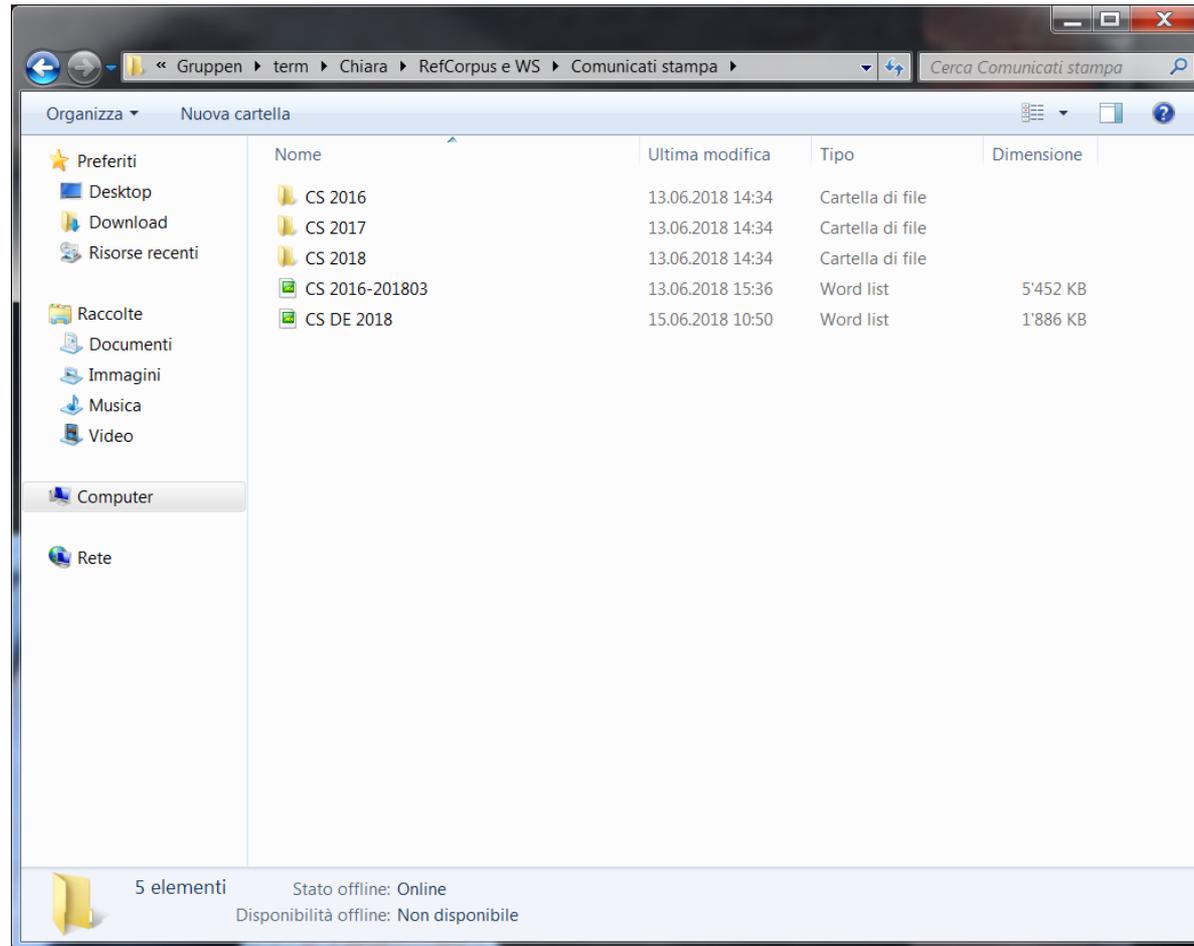
The main content area of the browser displays a directory listing for the URL `https://kav-search.bk.admin.ch/terminology/news/`. The listing is organized into a table with three columns: "Name", "Last modified", and "Size".

Name	Last modified	Size
Parent Directory		-
de/	2018-01-01 22:30	-
fr/	2018-01-01 22:30	-
it/	2018-01-01 22:30	-

At the bottom of the page, there is a copyright notice: © Gambazzi Consulting *Rocket propellant, nocino and metadata for the masses.*



# Corpus building





# Reference corpus > WL

SR 2016-08 WLlist

File Edit View Compute Settings Windows Help

N

	Overall	101_1	101_1	101_1	101_1	101_1	101_1	101_1	101_1	101_1	101_1	101_1	101_1	101_1	101_1	101_1	
text file																	
file size	1'340'109'952	403'115'299'558	296'242'298'700	305'731'318'000	318'000'361'000	364'000'379'000	373'000'400'000	401'000'402'000									
tokens (running words) in text	91'017'008	27'045'18'896	18'711'18'880	19'305'20'200	20'200'21'300	21'500'22'600	22'300'24'400	24'500'24'600									
tokens used for word list	91'017'008	27'045'18'896	18'711'18'880	19'305'20'200	20'200'21'300	21'500'22'600	22'300'24'400	24'500'24'600									
sum of entries																	
types (distinct words)	425'866	4'526	3'540	3'500	3'514	3'599	3'621	3'666	3'666	3'699	3'788	3'633	3'700	3'733	3'733	3'733	3'733
type/token ratio (TTR)	0.47	16.74	18.73	18.70	18.6	18.5	17.8	18.0	17.1	17.1	16.7	16.2	15.1	15.1	15.1	15.1	15.1
standardised TTR	35.73	44.87	45.20	45.33	45.1	45.6	44.8	44.9	45.1	45.2	44.7	44.4	43.3	43.4	43.3	43.3	43.3
STTR std.dev.	63.40	53.01	53.72	53.70	53.7	53.7	54.2	54.2	54.1	54.2	54.8	54.7	56.0	56.0	56.0	56.0	56.0
STTR basis	1'000	1'000	1'000	1'000	1'000	1'000	1'000	1'000	1'000	1'000	1'000	1'000	1'000	1'000	1'000	1'000	1'000
mean word length (in characters)	5.57	5.93	6.34	6.32	6.31	6.30	6.24	6.23	6.24	6.24	6.18	6.13	5.99	5.99	5.99	5.99	5.99
word length std.dev.	4.45	4.50	4.68	4.68	4.68	4.66	4.64	4.63	4.65	4.65	4.64	4.62	4.56	4.56	4.56	4.56	4.56
sentences	6'279'865	1'898	1'178	1'178	1'195	1'231	1'321	1'321	1'741	1'751	1'871	1'941	2'321	2'331	2'331	2'331	2'331
mean (in words)	14.69	14.25	16.04	15.89	15.7	15.7	15.3	15.2	12.2	12.2	12.0	11.4	10.5	10.5	10.5	10.5	10.5
std.dev.	202.29	23.27	16.62	17.36	17.3	17.1	17.2	16.8	15.7	15.4	15.3	15.2	14.6	14.6	14.6	14.6	14.6
paragraphs	11'239	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
mean (in words)	8'098.31	27'045	18'896	18'711	18'880	19'305	20'200	20'200	21'300	21'500	22'600	22'300	24'400	24'500	24'600	24'600	24'600
std.dev.	15'016.07																
headings																	
mean (in words)																	
std.dev.																	
sections	11'239	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
mean (in words)	8'098.31	27'045	18'896	18'711	18'880	19'305	20'200	20'200	21'300	21'500	22'600	22'300	24'400	24'500	24'600	24'600	24'600
std.dev.	15'016.07																
numbers removed																	
stoplist tokens removed	37'112'988																
stoplist types removed	532																
1-letter words	11'833'222	1'395	1'013	1'020	1'021	1'051	1'101	1'101	1'151	1'161	1'201	1'191	1'321	1'321	1'321	1'321	1'321
2-letter words	8'366'755	2'906	1'588	1'567	1'586	1'631	1'751	1'751	1'851	1'871	2'001	2'061	2'371	2'391	2'391	2'391	2'391

frequency alphabetical statistics filenames notes

77 entries Row 1 0% German T S text file



# Subcorpus > WL

N	Word	Freq.	% Texts	%person emmas Set
1	SCHWEIZ	392	0.65	113 60.1
2	BUNDESRAT	245	0.41	85 45.2
3	2016	209	0.35	96 51.0
4	SCHWEIZER	188	0.31	68 36.1
5	2015	179	0.30	66 35.1
6	PROZENT	117	0.19	36 19.1
7	APRIL	95	0.16	64 34.0
8	ZUSAMMENARBEIT	85	0.14	52 27.6
9	FRANKEN	84	0.14	31 16.4
10	MASSNAHMEN	81	0.13	40 21.2
11	1	80	0.13	47 25.0
12	MILLIONEN	71	0.12	28 14.8
13	RAHMEN	67	0.11	49 26.0
14	ENTWICKLUNG	66	0.11	42 22.3
15	PERSONEN	66	0.11	30 15.9
16	ZUDEM	66	0.11	53 28.1
17	INTERNATIONALEN	62	0.10	36 19.1
18	2014	61	0.10	33 17.5
19	BERICHT	58	0.10	19 10.1
20	WIRTSCHAFT	55	0.09	28 14.8
21	BUND	54	0.09	28 14.8
22	MÄRZ	54	0.09	30 15.9
23	UMSETZUNG	52	0.09	37 19.6
24	000	50	0.08	28 14.8
25	INSBESONDERE	50	0.08	42 22.3
26	EIDGENÖSSISCHEN	48	0.08	36 19.1
27	FORSCHUNG	48	0.08	20 10.6
28	STRATEGIE	48	0.08	18 9.57
29	ZIEL	48	0.08	35 18.6
30	MILLIARDEN	47	0.08	14 7.45



# Compare 2 WL > Results

N	Key word	Freq.	% Texts	RC	RC %	Log_L	Log_R	P emmas Set
1	SANS-PAPIERS	16	0.03	1	0	237.24	0.00	0.0000000000
2	GRUNDEINKOMMEN	15	0.02	1	0	222.41	0.00	0.0000000000
3	FRAGILEN	13	0.02	3	0	192.75	0.00	0.0000000000
4	GEWALTTÄTIGEM	12	0.02	3	0	177.93	0.00	0.0000000000
5	ASSELBORN	11	0.02	2	0	163.10	0.00	0.0000000000
6	SWISSNEX	11	0.02	2	0	163.10	0.00	0.0000000000
7	CLEANTECH	10	0.02	3	0	148.27	0.00	0.0000000000
8	PARMELIN	10	0.02	4	0	148.27	0.00	0.0000000000
9	DEAL	10	0.02	2	0	148.27	0.00	0.0000000000
10	STROMNETZE	9	0.01	1	0	133.44	0.00	0.0000000000
11	AUSSENMINISTERIN	9	0.01	4	0	133.44	0.00	0.0000000000
12	FRAGILITÄT	9	0.01	2	0	133.44	0.00	0.0000000000
13	STEINMEIER	9	0.01	2	0	133.44	0.00	0.0000000000
14	IDPS	9	0.01	3	0	133.44	0.00	0.0000000000
15	ENGAGIERT	8	0.01	6	0	118.62	0.00	0.0000000000
16	UNTERSTRICH	8	0.01	7	0	118.62	0.00	0.0000000000
17	BODLUV	8	0.01	2	0	118.62	0.00	0.0000000000
18	2.00	8	0.01	2	0	118.62	0.00	0.0000000000
19	RAPHAEL	8	0.01	3	0	118.62	0.00	0.0000000000
20	GASSER	8	0.01	1	0	118.62	0.00	0.0000000000
21	2017-2020	8	0.01	7	0	118.62	0.00	0.0000000000
22	VERBUNDPARTNER	8	0.01	1	0	118.62	0.00	0.0000000000
23	RETTUNGSTAUCHER	7	0.01	1	0	103.79	0.00	0.0000000000
24	VORJAHRESPERIODE	7	0.01	2	0	103.79	0.00	0.0000000000
25	RALPH	7	0.01	2	0	103.79	0.00	0.0000000000
26	SCHRAIVOGEL	7	0.01	1	0	103.79	0.00	0.0000000000
27	THEATERPREISE	7	0.01	1	0	103.79	0.00	0.0000000000
28	PIRACY	7	0.01	1	0	103.79	0.00	0.0000000000
29	MEHRKOSTENFAKTOR	7	0.01	1	0	103.79	0.00	0.0000000000
30	1,5	6		5	0	88.96	0.00	0.0000000000



# Really a term? WL > Concordance

N	Concordance	Set	Tag	Word	#	Ser	Ser	Par	Par	Hez	Hez	Sec	Sec	File	Date	%
1	die Initiative „Für ein bedingungsloses Grundeinkommen“ Die Volksinitiative				11	0	10	0	10			0	10	BR_2016-	2016.mag	2%
2	Volksinitiative „Für ein bedingungsloses Grundeinkommen“ will der ganzen				17	0	16	0	16			0	16	BR_2016-	2016.mag	3%
3	verlangt vom Bund, dass er ein Grundeinkommen einführt, das ohne				90	0	89	0	89			0	89	BR_2016-	2016.mag	17%
4	schlagen sie ein monatliches Grundeinkommen von 2500 Franken				138	0	13	0	13			0	13	BR_2016-	2016.mag	26%
5	Bundesrats hätte ein bedingungsloses Grundeinkommen einschneidende				166	0	16	0	16			0	16	BR_2016-	2016.mag	30%
6	der sozialen Sicherheit. Mit einem Grundeinkommen wäre es für				182	0	18	0	18			0	18	BR_2016-	2016.mag	34%
7	weniger oder nicht viel mehr als das Grundeinkommen verdienen, also für				209	0	20	0	20			0	20	BR_2016-	2016.mag	39%
8	Mittel nötigDas bedingungslose Grundeinkommen würde jährlich 208				258	0	25	0	25			0	25	BR_2016-	2016.mag	49%
9	der sozialen Sicherheit, die über das Grundeinkommen hinausgehen sowie				337	0	33	0	33			0	33	BR_2016-	2016.mag	63%
10	sozialen Sicherheit würde durch das Grundeinkommen weder vollständig				365	0	36	0	36			0	36	BR_2016-	2016.mag	69%
11	des Grundeinkommens erschwert.Das Grundeinkommen würde den sozialen				385	0	38	0	38			0	38	BR_2016-	2016.mag	73%
12	. Mit dem bedingungslosen Grundeinkommen erhielte die ganze				413	0	41	0	41			0	41	BR_2016-	2016.mag	79%
13	sozialen Zusammenhalt gefährden.Das Grundeinkommen stellt zudem zentrale				442	0	44	0	44			0	44	BR_2016-	2016.mag	84%
14	. Dieser würde aber mit dem Grundeinkommen an integrativer Kraft				467	0	46	0	46			0	46	BR_2016-	2016.mag	89%
15	die Gesellschaft. Ein bedingungsloses Grundeinkommen würde aber den				504	0	50	0	50			0	50	BR_2016-	2016.mag	96%



# Searching for definitions (defini\*)

The screenshot shows the Concord software interface. The main window displays a concordance search for 'defini\*'. The search results are highlighted in yellow and include terms like 'Definizione' and 'definire'. The interface includes a menu bar (File, Edit, View, Compute, Settings, Windows, Help), a toolbar, and a status bar at the bottom showing '146 entries' and 'Row 1'.

N	Concordance	Set Tag	Word	#	Ser	Ser	Par	Par	Hez	Hez	Sec	Sec	File
1	, si molto discusso a tal proposito. Definizione: il lungo cammino degli			141	0	14	0	14			0	14	AB_2012_
2	di valore aggiunto. Come vengono definiti i rifiuti alimentari? Per le mele			313	0	31	0	31			0	31	AB_2012_
3	viene mangiata? un rifiuto? Come definizione principale si imposta			339	0	33	0	33			0	33	AB_2012_
4	ammesse in entrata in un impianto, definita dalla Commissione ispettiva per le			1'98	0	1	0	1			0	1	Rifiuti orga
5	utilizzato in agricoltura sono state definite dall'Associazione Svizzera			1'78	0	1	0	1			0	1	Rifiuti orga
6	dell'industria e del commercio 5.1 Definizione			3'61	0	3	0	3			0	3	Rifiuti orga
7	(scheda S.8). Olii e grassi organici 3.1 Definizione Questa categoria			2'76	0	2	0	2			0	2	Rifiuti orga
8	positiva. Rifiuti organici di cucina 4.1 Definizione Questa			3'06	0	3	0	3			0	3	Rifiuti orga
9	Conclusioni 8 Legno naturale 30 8.1 Definizione 30 8.2 Produzione e			163	0	16	0	16			0	16	Rifiuti orga
10	7 Letame e colaticcio 27 7.1 Definizione 27 7.2 Produzione e			150	0	14	0	14			0	14	Rifiuti orga
11	G Rifiuti organici 9 Scarti vegetali 2.1 Definizione e base legale In questa			1'07	0	1	0	1			0	1	Rifiuti orga
12	digestato (solido e liquido) stata definita una lista positiva. Le			1'77	0	1	0	1			0	1	Rifiuti orga
13	che possono essere trattati sono definiti nella lista positiva 2. Numerosi			1'67	0	1	0	1			0	1	Rifiuti orga
14	sar possibile prendere decisioni e definire delle strategie d'intervento in			3'88	0	3	0	3			0	3	Rifiuti orga
15	/Comuni. Finanziamento Da definire. Tempistica Realizzazione a			6'81	0	6	0	6			0	6	Rifiuti orga
16	strategica. Fase successive da definire. Tempistica Termine della			6'28	0	6	0	6			0	6	Rifiuti orga
17	da Agridea e APF. Tempistica Da definire tra le parti di competenza. S.			7'27	0	7	0	7			0	7	Rifiuti orga
18	SPAAS-UPDA/SA. Finanziamento Da definire. Tempistica all'entrata in			7'43	0	7	0	7			0	7	Rifiuti orga
19	SPAAS/SA. Finanziamento Da definire. Tempistica all'entrata in			7'35	0	7	0	7			0	7	Rifiuti orga
20	sottoprodotti di origine animale. Art. 3 Definizioni 1 Sono			3'99	0	3	0	3			0	3	Rifiuti orga
21	pi circostanziata. Rifiuti animali 6.1 Definizione e base legale Questa			3'89	0	3	0	3			0	3	Rifiuti orga
22	organici 7 Letame e colaticcio 7.1 Definizione Questa categoria			4'35	0	4	0	4			0	4	Rifiuti orga
23	solonariamente reimpiantata. OS 2 Definire e implementare procedure di			5'22	0	5	0	5			0	5	Rifiuti orga



# Subordinate terms, synonyms & co. (\*abfälle)

The screenshot shows the Concord software interface with a concordance table. The table has columns for line number (N), concordance text, and various statistical fields (Set Tag, Word, #, Ser, Ser, Par, Par, Hez, Hez, Sec, Sec). The concordance text is highlighted in yellow, and certain words are color-coded: 'pflanzliche Rüstabfälle' in red, 'Gesamt mengen Lebensmittelabfälle' in green, 'Fleischabfälle' in blue, 'Bordabfälle' in purple, 'Speise-abfälle' in orange, and 'organischen Abfälle' in pink.

N	Concordance	Set Tag	Word	#	Ser	Ser	Par	Par	Hez	Hez	Sec	Sec	
1	Zusammensetzung 22'500 t pflanzliche Rüstabfälle (gerundete Werte) 250'000			1'89	0	1'	0	1'			0	1'	ÄK Ein M
2	Übersicht über die Gesamt mengen Lebensmittelabfälle aus dem Bereich			1'83	0	1'	0	1'			0	1'	ÄK Ein M
3	Kennzahl Zusammensetzung t Altöl t Fleischabfälle VTNP 1,2 Entsorgung t			1'74	0	1'	0	1'			0	1'	ÄK Ein M
4	bei der Berechnung der Menge Bordabfälle für die Schweiz jeweils nur			1'96	0	1'	0	1'			0	1'	ÄK Ein M
5	30% um Lebensmittelabfälle handelt. Bordabfälle werden jeweils am Zielort			1'94	0	1'	0	1'			0	1'	ÄK Ein M
6	-fälle, wovon es sich bei 30% um Lebensmittelabfälle handelt. Bordabfälle			1'94	0	1'	0	1'			0	1'	ÄK Ein M
7	. Tabelle 2 zeigt die Gesamt mengen Lebensmittelabfälle aus dem Bereich			1'72	0	1'	0	1'			0	1'	ÄK Ein M
8	in pflanzliche Rüstabfälle, Speise-abfälle bzw. Tellerreste sowie			1'44	0	1'	0	1'			0	1'	ÄK Ein M
9	der Gastronomie wurden in pflanzliche Rüstabfälle, Speise-abfälle bzw.			1'44	0	1'	0	1'			0	1'	ÄK Ein M
10	anfallen [8]. Die organischen Abfälle aus der Gastronomie wurden			1'43	0	1'	0	1'			0	1'	ÄK Ein M
11	resp. 0.1 t/d). Ein geringer Teil dieser Abfälle stammen aus dem			1'68	0	1'	0	1'			0	1'	ÄK Ein M
12	kg / Mahlzeit % pflanzliche Rüst-abfälle % Speisereste 5 RESULTATE			1'57	0	1'	0	1'			0	1'	ÄK Ein M
13	VTNP 85 % VTNP 3 1 % Altöl 2 % Fleischabfälle VTNP 1,2 Entsorgung 3			1'51	0	1'	0	1'			0	1'	ÄK Ein M
14	Hotelküchen ..... 12 Tabelle 5 Lebensmittelabfälle aus Catering			2'98	0	2'	0	2'			0	2'	ÄK Ein M
15	7 Abbildung 2 Entsorgungswege für Lebensmittelabfälle in der Schweiz .			2'92	0	2'	0	2'			0	2'	ÄK Ein M
16	, Transport) zu verzeichnen. Diese Abfälle sind mehrheitlich (95%) ver-			2'64	0	2'	0	2'			0	2'	ÄK Ein M
17	Einrichtungen ..... 14 Tabelle 8 Lebensmittelabfälle aus Schul- und			3'00	0	3'	0	3'			0	3'	ÄK Ein M
18	133 Tabelle 7 Lebensmittelabfälle aus medizinischen			2'99	0	2'	0	2'			0	2'	ÄK Ein M
19	13 Tabelle 6 Lebensmittelabfälle aus Kasernen.			2'99	0	2'	0	2'			0	2'	ÄK Ein M
20	ohne Restaurationsbetriebe, also ohne Lebensmittelabfälle, welche erst auf			2'62	0	2'	0	2'			0	2'	ÄK Ein M
21	Zusammensetzung 4'270 t pflanzliche Rüstabfälle 45'950 t Speisereste VTNP			2'36	0	2'	0	2'			0	2'	ÄK Ein M
22	'430 t pro Jahr (Tabelle 8). Tabelle 8 Lebensmittelabfälle aus Schul- und			2'34	0	2'	0	2'			0	2'	ÄK Ein M
23	- Zusammensetzung 270 t pflanzliche Rüstabfälle 2'910 t Speisereste VTNP			2'09	0	2'	0	2'			0	2'	ÄK Ein M



# ... resorted (L1)...

The screenshot shows the Concord software interface. The main window displays a concordance table with the following columns: N, Concordance, Set Tag, Word, #, Ser, Ser, Par, Par, Hez, Hez, Sec, Sec. The text in the concordance table is color-coded: green for 'pflanzliche Rüstabfälle', red for 'organische Abfälle', blue for 'Lebensmittelabfälle', and yellow for 'Bordabfälle'. The table shows 52 entries, with the first entry being 'Zusammensetzung 4'270 t pflanzliche Rüstabfälle 45'950 t Speisereste VTNP'. The interface also includes a menu bar (File, Edit, View, Compute, Settings, Windows, Help) and a status bar at the bottom showing '52 entries', 'Row 1', and '0%'.



# ...and R1

The screenshot shows the Concord software interface with a concordance table. The table has columns for 'N', 'Concordance', 'Set Tag Word', and various frequency counts. The text in the concordance column is highlighted in yellow, showing the phrase 'ste als vermeidbare Lebensmittelabfälle bezeichnet, während' across multiple lines. The interface includes a menu bar (File, Edit, View, Compute, Settings, Windows, Help) and a status bar at the bottom showing '52 entries', 'Row 2', and '0%'.

N	Concordance	Set Tag Word	#	Ser	Ser	Par	Par	Hez	Hez	Sec	Sec	File
1	Einrichtungen ..... 14 Tabelle 8 Lebensmittelabfälle aus Schul- und		3'008	0	3'0	3'				0	3'	ÄK Ein M
2	und Essensreste als vermeidbare Lebensmittelabfälle bezeichnet,		984	0	98	98				0	98	ÄK Ein M
3	in pflanzliche Rüstabfälle, Speise-abfälle bzw. Tellerreste sowie		1'441	0	1'0	1'				0	1'	ÄK Ein M
4	Kanton Bern berechnete für organische Abfälle aus der Gastronomie einen		1'378	0	1'0	1'				0	1'	ÄK Ein M
5	133 Tabelle 7 Lebensmittelabfälle aus medizinischen		2'998	0	2'0	2'				0	2'	ÄK Ein M
6	'430 t pro Jahr (Tabelle 8). Tabelle 8 Lebensmittelabfälle aus Schul- und		2'341	0	2'0	2'				0	2'	ÄK Ein M
7	, eine Verwertungspflicht von biogenen Abfälle (dazu gehören auch		638	0	63	63				0	63	ÄK Ein M
8	biogenen Abfälle (dazu gehören auch Lebensmittelabfälle) einzuführen, wenn		642	0	64	64				0	64	ÄK Ein M
9	genießbar wären. Unvermeidbare Lebensmittelabfälle entstehen		927	0	92	92				0	92	ÄK Ein M
10	oder Knochen als unvermeidbare Lebensmittelabfälle erfasst werden.		997	0	99	99				0	99	ÄK Ein M
11	der vorgesehenen Regelung müssen Lebensmittelabfälle der stofflichen		668	0	66	66				0	66	ÄK Ein M
12	ist in Anhang 4 TVA geregelt. Biogene Abfälle, die sich nicht zur stofflichen		725	0	72	72				0	72	ÄK Ein M
13	Postulats Pflicht zur Verwertung der Nahrungsmittelabfälle eingeführt werden		178	0	17	17				0	17	ÄK Ein M
14	Zusammensetzung 4'270 t pflanzliche Rüstabfälle 45'950 t Speisereste VTNP		2'361	0	2'0	2'				0	2'	ÄK Ein M
15	dem Detailhandel haben. Rund 1/8 der Lebensmittelabfälle (~50'000 t/a)		2'441	0	2'0	2'				0	2'	ÄK Ein M
16	und beinhalten im Wesentlichen Rüstabfälle aber auch essbare Speise-		938	0	93	93				0	93	ÄK Ein M
17	auf ca. Detailhandel Gastronomie Lebensmittelabfälle 100'280 t / a		2'581	0	2'0	2'				0	2'	ÄK Ein M
18	- Zusammensetzung 270 t pflanzliche Rüstabfälle 2'910 t Speisereste VTNP		2'091	0	2'0	2'				0	2'	ÄK Ein M
19	Lebensmittelabfälle 100'280 t / a Lebensmittelabfälle 289'650 t / a davon		2'581	0	2'0	2'				0	2'	ÄK Ein M
20	, gelten auch als vermeidbare Lebensmittelabfälle. Aufgrund dieser		968	0	96	96				0	96	ÄK Ein M
21	anfallen [8]. Die organischen Abfälle aus der Gastronomie wurden		1'431	0	1'0	1'				0	1'	ÄK Ein M
22	13 Tabelle 6 Lebensmittelabfälle aus Kasernen.		2'991	0	2'0	2'				0	2'	ÄK Ein M
23	Hotelküchen 12 Tabelle 5 Lebensmittelabfälle aus Catering		2'981	0	2'0	2'				0	2'	ÄK Ein M



# R1 > definition (sometimes)

The screenshot shows the Concord software interface. The main window displays a text document with the following content:

Lebensmittelabfälle werden in vermeidbare Lebensmittelabfälle und unvermeidbare Lebensmittelabfälle unterteilt. Bei den vermeidbaren Lebensmittelabfällen handelt es sich um Lebensmittelreste, die zum Zeitpunkt ihrer Entsorgung und bei rechtzeitiger Verwendung genießbar wären. Unvermeidbare Lebensmittelabfälle entstehen üblicherweise im Zuge der Speisenzubereitung und beinhalten im Wesentlichen Rüstabfälle aber auch essbare Speisereste ohne Konsumentennachfrage. Präferenz und Verteilverluste, die aufgrund von unterschiedlichen Gewohnheiten von Verbrauchern, wie z. B. Kochüberschuss, Ästhetik usw., entstehen, gelten auch als vermeidbare Lebensmittelabfälle. Aufgrund dieser Definitionen werden also unförmige Früchte, abgelaufene Produkte, Tellerreste und Essensreste als vermeidbare Lebensmittelabfälle bezeichnet, während Rüstabfälle, nicht essbare Teile, wie Schalen oder Knochen als unvermeidbare Lebensmittelabfälle erfasst werden.

4.2.2 Definition der Verursachergruppen Die Verursachergruppen Gastronomie und Detailhandel wurden für die bessere Übersicht

gorien

zentren" wurde im Verlauf der Arbeit Abstand genommen, da keine nachvollziehbaren Daten für diese Unterscheidung vorlagen. Die Daten zu der Unterkategorie „Frischmärkte" konnten nicht separat erfasst werden und wurden in die Kategorie Detailhandel integriert, da (i) wenig verlässliche Daten zu den anfallenden Abfallmengen bestanden und (ii) die absoluten Mengen potentieller **Abfälle** infolge Eigen-Entsorgung nicht massgeblich an den Gesamtmengen beteiligt waren.

4.2.3 Definition der Entsorgungswege Die aktuell vorherrschenden Entsorgungswege wurden für die Gesamtkategorien Detailhandel und Gastronomie quantifiziert. Folgende Entsorgungswege wurden dabei berücksichtigt: Kehrichtverbrennungsanlagen (KVA), Kompostierung, Vergärung, Biodieselproduktion, Verfütterung und Tafel. Die Kategorie Tafel beinhaltet wohltätige Vereine, welche übriggebliebene oder aussortierte Lebensmittel bei Grossverteilern und Gastronomiebetrieben abholen und an bedürftige Menschen verteilen. Abbildung 2 gibt einen Überblick über die berücksichtigten Entsorgungswege für Lebensmittelabfälle in der Schweiz.

Abbildung 2 Entsorgungswege für Lebensmittelabfälle in der Schweiz Bei der Analyse der Massenströme aus den Bereichen Gastronomie und Detailhandel wurden folgende Annahmen getroffen: 1. Es wird davon ausgegangen, dass aus dem Detailhandel in massgeblichen Mengen in der KVA lediglich die Abfallfraktion aus Metzgereibetrieben verwertet wird, die als tieri

2.

wurde

esatz

The interface also shows a menu bar (File, Edit, View, Compute, Settings, Windows, Help) and a status bar at the bottom with the following information:

concordance	collocates	plot	patterns	clusters	timeline	filenames	source text	notes
52 entries	11 found	0%					T S	< >



# Clusters (Concordance) > MW terms?

The screenshot shows the Concord software interface with a concordance table. The table has columns for N, Cluster, Freq., Set, Length, and Related. The first entry is highlighted in yellow.

N	Cluster	Freq.	Set	Length	Related
1	DEGLI SCARTI VEGETALI	24	3		DEGLI SCARTI VEGETALI NEC
2	DI SCARTI VEGETALI	18	3		QUANTITATIVI DI SCARTI VEG
3	AGGIORNAMENTO DEL PGR S	12	3		AGGIORNAMENTO DEL PGR S
4	AGGIORNAMENTO DEL PGR S	12	4		AGGIORNAMENTO DEL PGR (
5	DEL PGR S	12	3		AGGIORNAMENTO DEL PGR S
6	QUANTITATIVI DI SCARTI VEGETALI	9	4		DI SCARTI VEGETALI (18),QU
7	SCARTI DI CUCINA	9	3		ALIMENTARI E SCARTI DI CUC
8	SCARTI DI DISTILLAZIONE	9	3		
9	QUANTITATIVI DI SCARTI	9	3		QUANTITATIVI DI SCARTI VEG
10	E SCARTI DI	7	3		RESTI ALIMENTARI E SCARTI
11	SCARTI VEGETALI RACCOLTI	7	3		QUANTITATIVI DI SCARTI VEG
12	OLI E GRASSI	6	3		SCARTI ANIMALI OLI E GRASS
13	SCARTI ORGANICI ANNUA SEGUITADAGLI SCARTI	6	5		SCARTI ORGANICI ANNUA SE
14	NEGLI IMPIANTI DI	6	3		VEGETALI NEGLI IMPIANTI DI
15	SCARTI ORGANICI ANNUA SEGUITADAGLI	6	4		SCARTI ORGANICI ANNUA SE
16	SCARTI DI LAVORAZIONE RAMAGLIA	6	4		SCARTI DI LAVORAZIONE RAI
17	SCARTI DI LAVORAZIONE RAMAGLIA E	6	5		SCARTI DI LAVORAZIONE RAI
18	SCARTI ORGANICI ANNUA	6	3		SCARTI ORGANICI ANNUA SE
19	SCARTI VEGETALI E	6	3		SCARTI VEGETALI E DI SCAR
20	SCARTI VEGETALI IMPIANTI	6	3		SCARTI VEGETALI IMPIANTI D
21	SCARTI VEGETALI IMPIANTI DI	6	4		SCARTI VEGETALI IMPIANTI (
22	SCARTI VEGETALI LETAME	6	3		SCARTI VEGETALI LETAME E
23	SCARTI VEGETALI E DI SCARTI	6	5		SCARTI VEGETALI E (6),SCAR
24	QUANTITATIVI DI SCARTI VEGETALI RACCOLTI	6	5		DI SCARTI VEGETALI (18),QU

concordance collocates plot patterns clusters timeline filenames source text notes  
160 entries Row 1 0% T S < > 1



# Patterns

Concord

File Edit View Compute Settings Windows Help

N	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5	
1		DI	DI	DI	QUANT	DEGLI	SCART	VEGET	DISTILI	DI	DI	SCART
2		SCARTI	DEL	PGR	ALIMEN	DI		DI	CUCIN	IMPIAN	SCART	COMP
3	AGGIORNAMENTO	SCART	2.2-1	PRODU	GLI		ORGAN	DI	DELLA	PER	OBIE	T
4	PIANIFICAZIONE	IMPIAN	IMPIAN	COMP	600		DELLA	PREPA	CANTC	COMU	VEGET	
5		DEGLI	ANIMA	0.3	DI	598		DELL	RACCC	DAI	FRUTT	800
6		2.2-2	PER	VEGET	GESTI	CON		ORGAN	COLAT	DEGLI	DAI	CIBO
7		VEGETALI	RIFIUT	RESTI	FORNIT	POPOL		VENGC	NEGLI	0.3	COMP	DEGLI
8		LEGENDA	TABELI	DEI	168	ICTR		SONOI	IMPIAN	PER	STATO	ORGAN
9		LA	FIGUR	ORGAN	NEICAF	ORGAN		VEGET	NON	INDUST	MODO	NON
10	NECESSARI	EVOLU	2011	LETAM	QUANT			VEGET	VALOR	ARROT	STATE	LIMITA
11	PER	AZIENE	IL	ORGAN	TRATT			INUTILI	INDUST	IL	NEL	INDUST
12	OPPUREIL	DELL	DELLA	NEL	TRATT			ANIMA	OTTEN	SONO	METAN	MACIN
13	PRINCIPALI	PRIMA	CA	TIPO	UTILIZZ			ALIMEN	OLI	ANCHE	GESTI	NO
14	TRATTA	MONIT	AGRIC	TERMC	SEGUI			AGRIC	PER	OTTIEN	TABELI	NATUR
15	TERMICO	QUEST	LO	TRATT	QUEST			CUCIN	SERVIZ	OTTEN	IMPIAN	SIA
16	UNA	NUTRIF	RIFIUT	TRATT	RESTC			INDUST	SUDDI	PULIZI	GRASS	SERVIZ
17	UFV	MONIT	PER	TABELI	RISTOI			IN	MONIT	PROVE	RIDUZI	TRATT
18	RESTI	ANNUA	OLI	RIFIUT	GRASS			IMPIAN	NEI	SUDDI	PERIQ	VIA
19	PRINCIPALIFONTI	SIMILE	OPERA	RAMAC	800				OBIE	SEGUI	SEGUI	VEDI
20	TASSE	1.2	ULTER	SMALT	9.3.2				SI	TICINO	SMALT	PESCE
21	SEPARATA	VEGET	PROVE	SMALT	732				PRODU	VERDU	SMALT	OTTIM
22		9.3.2	SECON	QUEST	AZIENE	2.2-2			TENUT	UN	NUTRIF	PRESS
23	5DELLAPRODUZIONE	2003	PICCO	ATTUA	4DELLA				PROVE	SEGES	NORM	REDDI
24	CAPITOL	10.2	SITUAZ	CON	ALLADI				PAPP	PESIDI	SOTT	PRIVAT

concordance collocates plot patterns clusters timeline filenames source text notes

160 entries Row 1 0% T S < >



# Exploring with ConcGrams

The screenshot shows the ConcGrams application window titled 'index\_file\_spreco.base\_pairs'. The interface is split into two main panes. The left pane displays a list of words and their frequencies, with 'RIFIUTI' selected. The right pane shows a detailed breakdown of the selected word into various concgrams, including their frequencies and counts.

Word	Freq
DI	539
<b>RIFIUTI</b>	<b>400</b>
E	239
IN	205
ORGANI...	193
AGGIO...	182
IMPIANTI	169
DEL	164
A	155
DEI	154
PER	154
CAPITO...	112
G	112
NOVEM...	109
LA	93
PRODU...	87
SCARTI	76
SMALTI...	75
FINANZI...	69
PRESE...	67

Concgram	Freq	Details
<b>RIFIUTI</b>		
PER-RIFIUTI-DI	(+6)	1
RIFIUTI-ALIMENTARI	(+46)	
RIFIUTI-ORGANICI	(+125)	1
AGGIORNAMENTO-PGR-RIFIUTI	(+6)	1
CANTONALE-PER-RIFIUTI		2
DEL-COMMERCIO-RIFIUTI		1
DEL-TICINO-RIFIUTI	(+1)	1
FIGURA-DI-RIFIUTI	(+4)	4
I-RIFIUTI-DI		2
IN-CASO-DIFERMO-RIFIUTI-IMPIANTI		1
IN-I-RIFIUTI-SVIZZERA		1
IN-MODO-I-RIFIUTI		2
IN-TICINO-SITUAZIONE-RIFIUTI		1
LA-DI-RIFIUTI		2
PER-DI-RIFIUTI	(+3)	4
PER-RACCOLTA-RIFIUTI-ANIMALI	(+1)	1
PRODUZIONE-ANNUA-DI-RIFIUTI-SUDDIVISA		4
RACCOLTA-E-RIFIUTI		1



# Exploring with ConcGrams

The screenshot displays the ConcGrams software interface. The main window is titled "index\_file\_spreco.base\_pairs". The interface is divided into several panes:

- Left Pane:** A table showing word frequencies. The word "SCARTI" is highlighted in blue.
- Right Pane (Details):** A tree view showing the breakdown of "SCARTI" into various concgrams, such as "SCARTI-IMPIANTI-DI" and "SCARTI-VEGETALI".

Word	Freq
DI	539
RIFIUTI	400
E	239
IN	205
ORGANI...	193
AGGIO...	182
IMPIANTI	169
DEL	164
A	155
DEI	154
PER	154
CAPITO...	112
G	112
NOVEM...	109
LA	93
PRODU...	87
<b>SCARTI</b>	<b>76</b>
SMALTI...	75
FINANZI...	69
PRESE...	67
ALIMEN...	66
GESTIO...	60
DELL	58
COMPO...	57
SONO	57
VEGETAL	56
VIGORE	55

Concgram	Freq.
SCARTI	
SCARTI-IMPIANTI-DI	(+6) 1
DEGLI-SCARTI-DI	1
GLI-SCARTI-DI	3
IMPIANTI-A-LEGNA-SCARTI	1
IMPIANTI-DI-SCARTI	(+4) 2
SCARTI-DELLA-DI	2
SCARTI-DI-CUCINA	(+1) 1
SCARTI-DI-CUCINA-PER	1
SCARTI-DI-DELLA	(+2) 1
SCARTI-E-DI	(+3) 1
SCARTI-DI-T	(+1) 2
SCARTI-E-CUCINA	1
SCARTI-ORGANICI-T	1
SCARTI-VEGETALI	(+42) 2
SCARTI-VEGETALI-E	(+1) 5
SCARTI-VEGETALI-IMPIANTI-DI-COMPOSTAGGIO-OB...	1
SCARTI-VEGETALI-IMPIANTI-DI-OBIETTIVO	1
SCARTI-VEGETALI-IMPIANTI-DI-COMPOSTAGGIO	2
SCARTI-VEGETALI-E-DI-ORGANICI	1
SCARTI-VEGETALI-IMPIANTI-DI	1
SCARTI-VEGETALI-PER-COMPOSTAGGIO	1
SCARTI-VEGETALI-T	2
SCARTI-VEGETALI-PER	(+1) 3
SCARTI-VEGETALI-OBIETTIVO	(+1) 2
SCARTI-VEGETALI-OBIETTIVO-IMPIANTI	1
SCARTI-VEGETALI	5