

Qu'est-ce qu'une machine linguistique ? Épistémologie de la technique et linguistique instrumentée

Pierre-Yves Modicom
Université Bordeaux Montaigne

ANNEXES À LA PUBLICATION

Compte tenu des contraintes du support, le texte ne permet que la reproduction en nombre limité d'images noir et blanc et au format vignette. Je redonne ici l'ensemble des figures « inédites » mentionnées en note, puis je donne une version en couleur et agrandie des figures N & B du texte. Toutes les images sont des captures d'écran réalisées par moi et sont sous licence CC BY-NC-SA.

1. Figures inédites

Figure 1 : inventaire des occurrences de la forme *denke* dans le corpus DRKorp (Modicom 2016) , résultats sur AntConc.

The screenshot shows the AntConc 3.4.4w (Windows) 2014 interface. The search term is 'denke' and the total number of hits is 90. The results are displayed in a table with columns for Hit, KWIC, and File. The KWIC column shows the context of the word 'denke' in various sentences, such as 'zurückgezogen hat und so weiter. Ich denke auch, dass es der Opposition letztendlich für den Einsatz von Chemiewaffen. Ich denke, dass es deshalb auch insgesamt Syrien auch in Moskau? Rosiny; Also ich denke, der Schlüssel, der liegt bei allen bei allen auf allen Seiten. Ich denke, das große Problem ist, dass sowohl eine sehr komplexe Gemengelage und deshalb denke ich kann es nur eine politische USA und Russland notwendig, und ich denke auch, dass es notwendig ist, den zurückgezogen hat und so weiter. Ich denke auch, dass es der Opposition letztendlich für den Einsatz von Chemiewaffen. Ich denke, dass es deshalb auch insgesamt Syrien auch in Moskau? Rosiny; Also ich denke, der Schlüssel, der liegt bei allen bei allen auf allen Seiten. Ich denke, das große Problem ist, dass sowohl eine sehr komplexe Gemengelage und deshalb denke ich kann es nur eine politische USA und Russland notwendig, und ich denke auch, dass es notwendig ist, den sich deshalb daran orientieren. Und ich denke, es gibt auch, oder es gibt die Kinder anderer Leute, ne? Ich denke, das ist heute auch ganz wichtig. Personen mit Sekundarabschluss oder Lehre. Ich denke auch das ist ein Maß, das auf alle Fälle studieren? Schleicher: Ich denke, man muss sich von dieser Vorstellung könnte man das ändern? Schleicher: Ich denke, wenn es immer mehr junge Menschen Tief, aber er hat seine Möglichkeiten denke ich genutzt. Dazu trägt auch die darf man nicht überschätzen. Aber ich denke schon, im Nachhinein werden die Grünen auch Rücksicht nehmen und ä ich denke aber, dass letzten Endes eben doch der Weise entscheiden würden. Aber ich denke, wir sind ganz klar auf der mit einheimischen ä Leuten. Und ich denke, jetzt würde es sich auch für ganz legitim und wichtig, aber ich denke, unsere Aufgabe wird sein, vielleicht auch - Bereichen tun. Gnath: Genau. Also ich denke, dass diese Aufmerksamkeit und das Engagemen liegt das denn? Erler: Ja, ich denke, das liegt daran, dass dieser diese durchaus positiv stimmen könnte? Polenz: Das denke ich schon, allerdings muss man auch ist, sich nicht äußern dürfen. Ich denke, der Sicherheitsrat sollte den International



Figure 2 : Inventaire des mots du corpus DRKorp classés par fréquence, résultats sur AntConc.

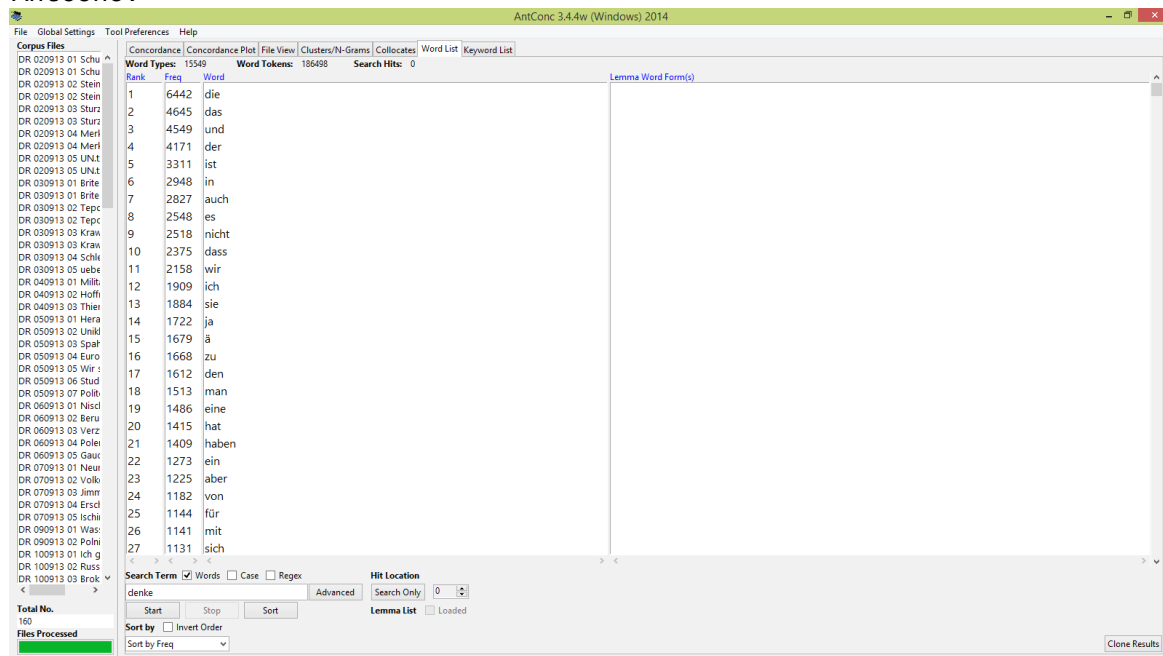


Figure 3 : Collocations les plus fréquentes pour la forme *denke* dans le corpus DRKorp, résultats sur AntConc.

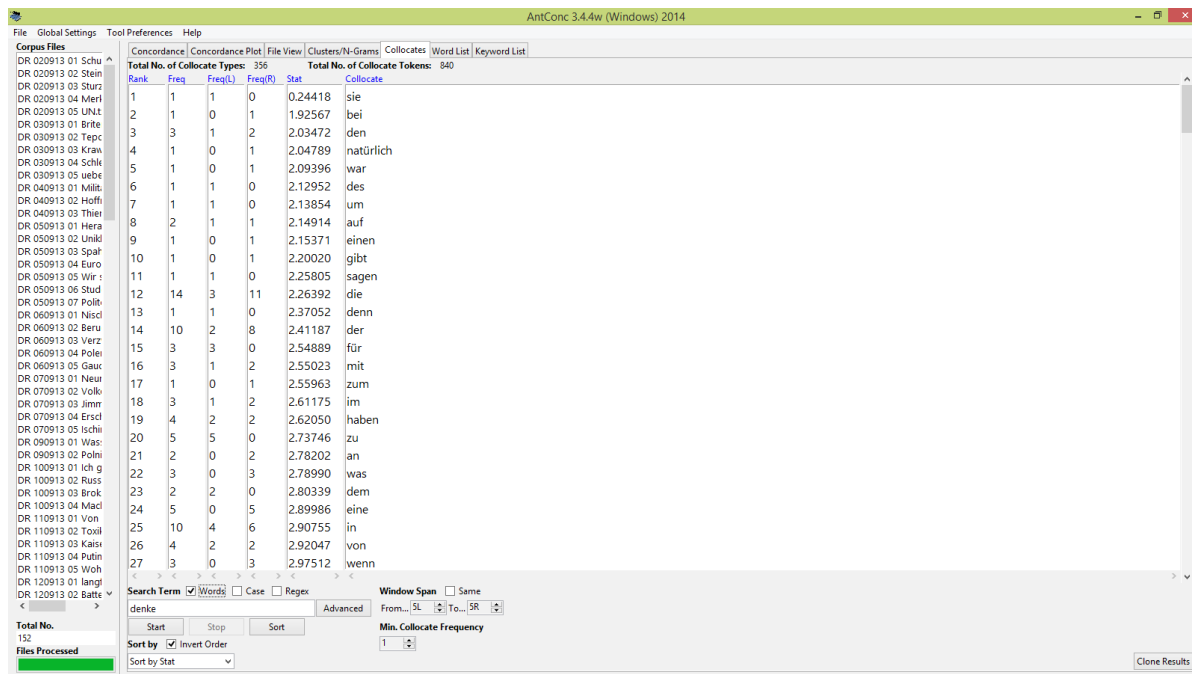


Figure 4 : inventaire des occurrences d'une forme du lemme *denken* dans le corpus DRKorp, résultats sur TXM.

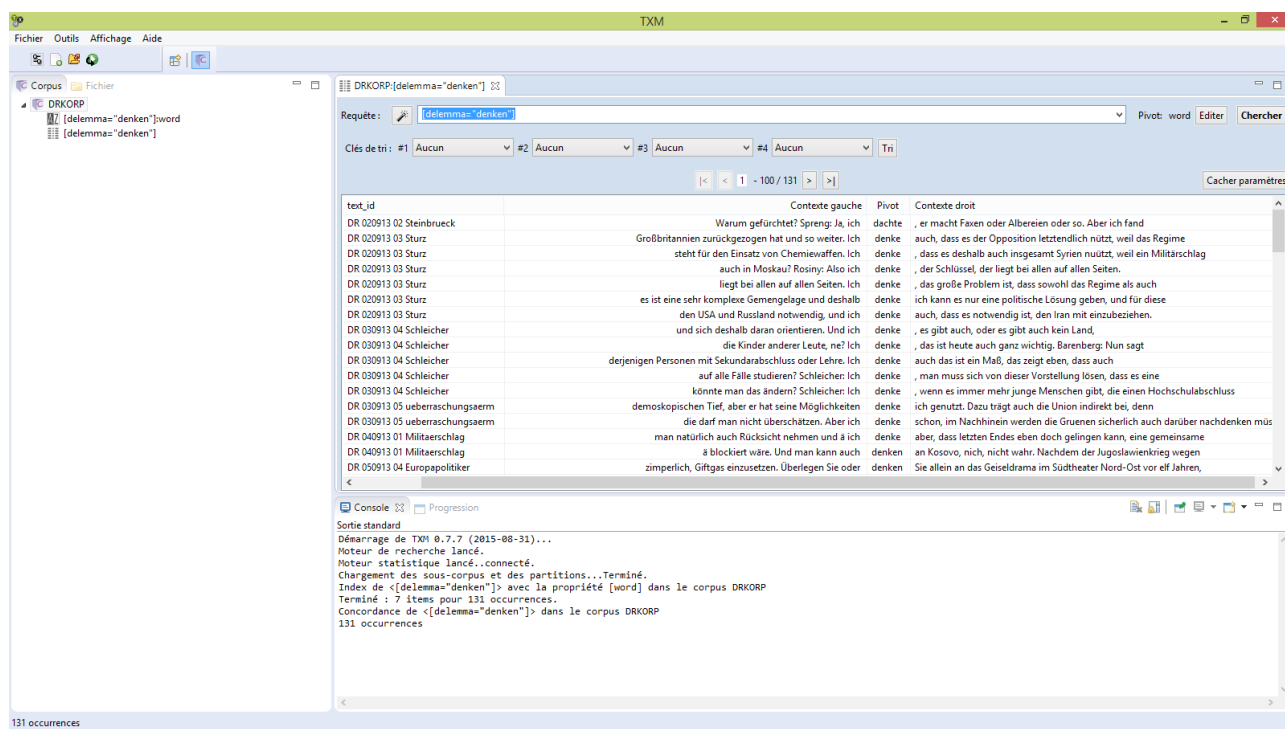


Figure 5 : Fréquence des collocations dans le corpus DRKorp pour le lemme *denken*. Résultats sur TXM.

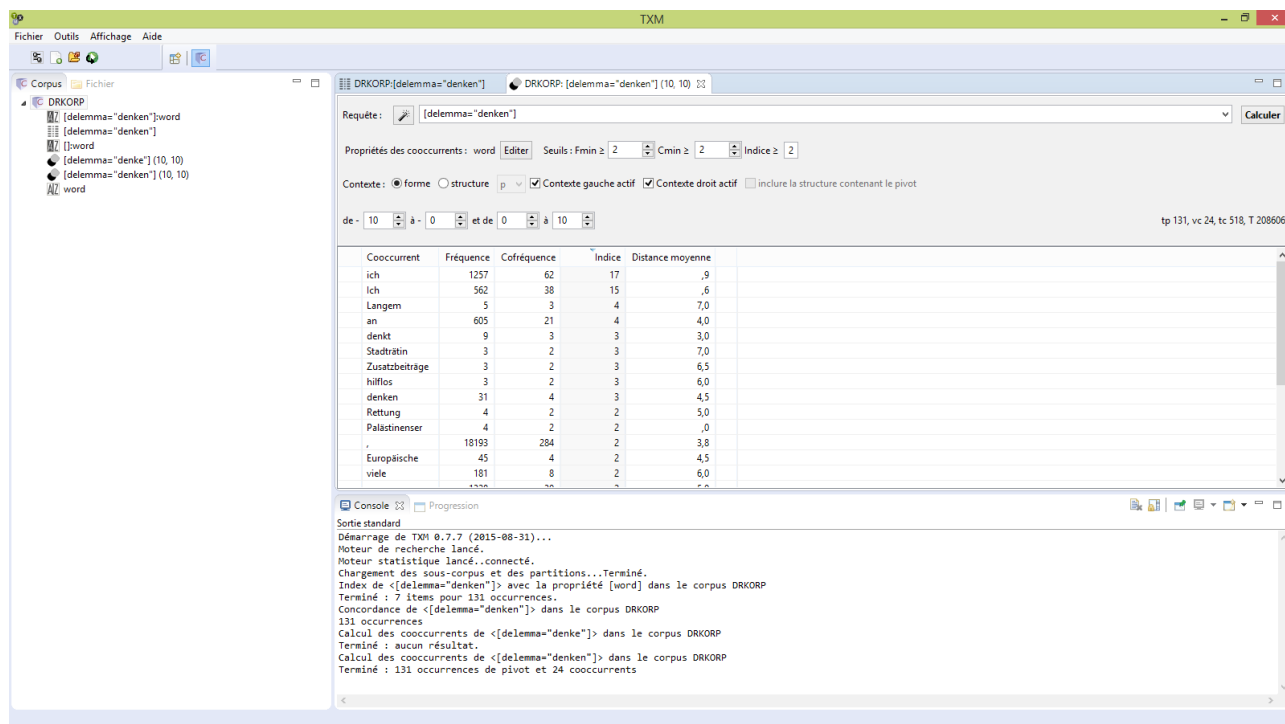


Figure 6 : Inventaire des exemples de séquence comprenant une forme du lemme *Germanist* suivie d'une forme du lemme *Tagung* avec un écart compris entre 0 et 8 mots dans le Deutsches Referenzkorpus (DEREKO), Archive W (aperçu des résultats de la requête sur COSMAS 2, 7 juin 2017)

The screenshot shows the COSMAS II web interface. The search query is `*Germanist +w1:9 *Tagung`. The results are sorted by KWIC (unsorted) and show 10 hits. The first result is: "N93/AUG.28465 ...che Ausrichtung der akademischen Welt zu einem bestimmten Thema", erklärt der Salzburger Germanist Ulrich Müller, einer der wissenschaftlichen Leiter der Tagung." Other results include mentions of Daniel Wilson, Per Ohrgaard, Dieter Borchmeyer, Wilhelm Solms, Leonhard M. Fiedler, Claudio Magris, Wolf gang Thierse, Karl-Josef Kuschel, and Michael Wögerbauer.

2. Version en couleur des figures N&B

Figure I: Liste de cooccurents pour une requête « verbe modal conjugué suivi d'un adverbe avec un écart compris entre 0 et 5 mots » sur TXM, corpus DRKorp (Modicom 2016).

The screenshot shows the TXM software interface. The query is `[depos="VMFIN"]][0..5][depos="ADV"]`. The results are displayed in a table with the following columns: Cooccurrent, Fréquence, Cofréquence, Indice, and Distance moyenne.

Cooccurrent	Fréquence	Cofréquence	Indice	Distance moyenne
sein	313	123	26	4,7
sagen	435	145	22	4,2
werden	685	184	16	3,9
man	1328	299	14	5,5
nicht	2377	468	11	5,5
Man	117	47	10	2,1
vorstellen	35	18	6	3,4
machen	243	65	6	5,2
passieren	21	12	5	3,3
da	804	161	4	5,8
weitergehen	13	9	4	3,0
tun	170	46	4	5,2
wer	91	29	4	6,6
fragen	38	16	4	4,1
nachdenken	9	7	4	6,1
Aber	463	98	4	5,9
Vetomächte	5	5	4	6,2
was	751	147	4	5,8
nationale	13	8	3	9,4
mal	466	97	3	5,2
geben	124	34	3	3,9
Und	818	155	3	5,3
haben	164	41	2	4,7



Figure II: Aperçu d'une fenêtre de travail sur le DDD, avec les annotations d'un énoncé du Tatian.

The screenshot shows the DDD interface with a search for 'uuoort'. The results table displays morphological information for the word 'uuoort' in the context of the Tatian text. The table includes columns for the word and its various grammatical forms and classes.

edition	In	anaginne	uuas	uuoort	inti	thaz	uuoort	uuas	mit
text	In	anaginne	uuas	uuoort	inti	thaz	uuoort	uuas	mit
lemma	in	anagin	wesan	wort	inti	der	wort	wesan	mit
postLemma	AP	NA	VA	NA	KO	DD	NA	VA	AP
pos	APPR	NA	VVFIN	NA	KON	DDA	NA	VAFIN	APPR
inflectionClassLemma		A_NEUT	ST5	A_NEUT			A_NEUT	ST5	
inflectionClass		A_NEUT	ST5	A_NEUT			A_NEUT	ST5	
inflection		SG_DAT	IND_PAST_SG_3	SG_NOM		NEUT_SG_NOM	SG_NOM	IND_PAST_SG_1	
lang	goh	goh	goh	goh	goh	goh	goh	goh	goh
clause		CF_U_M				CF_I_M			
document	Tatian								
chapter	1								
translation	in, an	Anfang	sein, werden	Wort	und, auch	der, die, das	Wort	sein	mit, zusammen mit, bei, unter

Figure III: Table des cooccurrences pour une requête « verbe modal conjugué suivi d'un adverbe, avec un écart compris entre 0 et 5 mots » pour TreeTagger sur l'archive T du DeReKo, 7 juin 2017.

The screenshot shows the COSMAS II interface with a co-occurrence analysis for the query 'MORPH(VRB fin m) /+w1:6 MORPH(ADV)'. The table lists various modal verbs and their associated adverbs with their respective frequencies and syntactic patterns.

#	LLR	kumul.	Häufig	Kooccurrenzen	syntagmatische Muster
1	279726	198307	198307	man	26% kann [...] man
2	265194	541049	342742	werden	7% werden [...] kann
3	235087	896695	356646	nicht	13% kann [...] nicht
4	174743	1035253	138558	sein	9% kann [...] sein
5	99281	1035282	29	Man gespannt	82% Man darf [also]auf das [...] gespannt
			1070957	Man	29% Man [...] kann
6	72821	1147461	76504	ich	17% kann [...] ich
7	72446	1237361	89900	wir	18% müssen [...] wir
8	44466	1237434	73	sagen nein	23% Wer kann da schon nein sagen
			1245190	sagen	21% Ich kann [nur ...] sagen
9	43555	1245431	241	Wir euer	100% Wir dürfen euer Bild nur
			1307950	Wir	31% Wir [...] wollen
10	19334	1346924	38974	ich	22% Ich [...] kann nur ...
11	19091	1346952	28	passieren Besseres	50% konnte gar nichts Besseres passieren als
			1347002	passieren Schlimmste	50% Das/das Schlimmste was einem ... passieren kann
			1353417	passieren	34% kann [...] passieren dass

