Language Sketching

ragos Ciobanu, from the translation department of the University of Leeds, wrote to me a few weeks ago:

The reason I'm writing is to ask whether you've had a chance to play with Sketch Engine (see sketchengine. co.uk). In Leeds, we've been using it in our corpus linguistics work a lot. It's got brilliant features, from the terabytes of super useful multilingual data which it already comes with, to features for term extraction, specialized corpus building, thesaurus, collocations, and tons more! It's really, really cool and I'm only writing to you because the translators I know who have been playing with it also like it a lot.

Not sure whether you could tell, but Dragos really likes Sketch Engine. And in a way, I could stop this column right here, because he already said it all-sort of.

After spending some time looking at Sketch Engine, I felt embarrassed that I hadn't known more about it. As Dragos said, it's really, really cool. It's also a monster of a tool (size-wise) and it's not particularly easy to navigate when you first encounter it. (According to Ondřej Matuška of the Sketch Engine team, one of the areas they're trying to focus on in the immediate future is to make the product more user-friendly.)

But first, what exactly is Sketch Engine and what does it do?

It's a corpus tool developed by the Czech company Lexical Computing Limited. Lexical Computing was originally founded in 2003 by the late Brit Adam Kilgarriff and Pavel Rychlý, a professor

at Masaryk University in Brno. The idea of corpus tools, and this corpus tool in particular, is to find how language behaves based on large collections of data. For this purpose, Sketch Engine built corpora in more than 80 languages (as well as "timestamped" corpora in a slightly different set of 18 languages for the purpose of comparing word usage over time). The sizes of the corpora differ widely (from just a few million words in Maori to more than 800 billion in English), and they are available for a number of analysis purposes for any paying trial user. (The annual subscription price is 100 euros for non-academic users, with the trial period ending after 30 days.)

The analyses you can do on these corpora with Sketch Engine include the following:



	agh."	DOUD	modified by "breakthreak	web.		verbs with "brea	Attrough as	object	verbs with "break	chrough" as a	ubject	"breakthrough" and/or				
91,97	44.92			60,452	29.53		60,601	29.60		21,152	10.33		19,014	9.29		
technological + 4.47	8.47	disco	ery +	1,489	7.12	achieve +	3.015	\$.70	bleed +	263	6.93	breakdown +	213	6.79		
bechnological breakth		br	eakthrough discovery	_		represent +	1.584	4.79	breakthrough b			breakdowns and bre		dis .		
scientific + 1.11	6.86	bleed	ing +	643	7.09	herald	22	4.65	parent	7.6	5.78	discovery +	242	6.21		
scientific breakthroug			eakthrough bleeding	-		pioneer +	122	4.52	Breakthrough i			advance +	246	5.86		
nevolutionary + 62			ation +	1.276	6.30		91	4,17	anti-age	17	4.60	invention +	118	5.70		
a revolutionary break			eakthrough innovation	- AND		yield +	157	4.12	breakthrough a			innovation +	550	5.51		
major + 7.90			elog +	4.481	5.17	mark +	292	4.05	occur +	727	4.15	advancement +	106	5.48		
a major breakthrough			eakthrough technology	Dallas.		game-changing	34	4.07	breakthrough o			breakthroughs and a				
diplomatic + 30		seizu		158	5.33			16	Dave	25	3.95	deliverance	33	5.22		
a diplomatic breakthy			eakthrough seizures			life-changing	38	3,97	breakthrough p			miracle	66	5.14		
conceptual + 25		CYUA		75	5.14				patent	14		epiphany	22	4.86		
conceptual breakthro			eakthrough evulation	-		Pal	45	3,90	a breakthrough			oure	57	4.69		
significant + 2.09		collab		83	4.97	astound	52		happen +	470	2.54	setback	31	4.65		
a significant breakthy			eakthrough Collaborative		-	astounding br			breakthroughs			miestone	21	4.49		
technology + 1.13			nation +	161	4.94	signal	22	3.80	come +	2.839	2.00	insight +	160	4.38	1	
technology breakthro			eakthrough therapy desig			commercialize	25	3.80	breakthrough o			insights and breakth			1	
dramatic + 42		Inver		146	4.74	exploit	22	3.75	revolutionize	14	3.94	transformation	64	4.38	1	
a dramatic breakthro				149	-	announce +	508	3,73	anoint	9		shift	66	4.33		
decisive + 14		4+	eakthrough invention	101	4.63		\$33	3.45	ecourt	11	3.20	revelation	10	4.33	1	
a decisive breakthrou			eakthrough DJ / Produce			witness +	127	1.41	breakthrough			sis	15	4.26		
medical + 2.58		pain 4			4.62		112	3.30	hit +		3.04	a breakthrough , sis		4.49		
			r breakthrough pain	1.020	1.74		50	1.21	breakthrough I	207				4.72		
medical breakthrough				747	4.40	precede		2.62			2.00	realization	24			
remarkable + 🔛		albur		762	4.49				buoy		2.99	victory	57	4.21		
a remarkable breakth			eakthrough album,		000000	promise	22	3.89	enable	22	2.45	lesp	23	4.19		
	4.47															
exciting + 55		forma	- Aller	122	4.15	catal) pr	12	1.19	ploreer breakthrough	10	2.42	achievement	-	1.91		
		forma	attective predicates of	-		Catalype Dreakthrough" is a	Contra a	0.70	possessors of 2	ser.	1.38		-	1.93	3.223	2.58
- Contraction		12.93	- sou	(Dreakthe	1.17	- and the second		~	persenters of 2	reakthrough	1	Noon	-	1.91	5.275 21	2.58
prepositional phrases	74.968	~	adjective predicates of	7 breakth 2,400 123	1.17 6.22	breakthrauch" is a	L411	0.70	breakthrough	reskithcough" 2,820	1.38		-	1.93		
Drepositional phrases	74.968 25.477	12.93	adjective predicates of	7 breakth 2,400 123	1.17 6.22	breakthrough' is a neorporation	1.433 2 20	0.70	breakthrough	reaktbrough 2.820 12	1.38	u. Is a "breakthrough"			21	6,40
Dresositional phrases Dresositional phrases Dresositional phrases of "breakthrough" on to "breakthrough"	74,968 25,477 9,155	12.93	adjective predicates of single + breakthrough single	C Dreakthe 2.400 132 54	5.41	breakthrough' is a neorporation nvention	1.433 2 20	0.70	possessors of 2 science ndp beatles	reakthrough 2.820 12 2	1.38 6.11 6.02 5.35	discovery bk			21	6,40
prepositional phrases	74.968 25.477 9.155 5.251	12.93 4.47 2.56	stiective predicates of sings + breakthrough single creative	C Dreakthe 2.400 132 54	5.41	breakthrough is a neorporation invention breakthrough wi	Len 1.433 2 30 as the invention 21	0.70 6.33 6.21	possessors of 2 science ndp beatles	reakthrough 2.820 12 9 10 90	1.38 6.11 6.02 5.35	discovery bk Cynengy TK is a bre			21 12	6.40
Dresoltional phrases	74,968 25,477 9,153 5,251 4,519	12.93 4.47 2.56 2.21	adiective predicates of singe + breakthrough single breakthrough creat	2.400 123 54 122	5.41 5.29	corporation incorporation invention breakthrough wi discovery	Len 1.433 2 30 as the invention 21	0.70 6.33 6.21	Desketsors of a science ndp beatles band the bands br	reakthrough 2.820 12 9 10 90	1.38 6.11 6.02 5.35 4.93	usering discovery St Cytergy TK is a bre technology +			21 13 123	6.10 6.10
Presettional phrases Dresettional phrases Dresettional? If of "breakthrough" 	74,968 25,477 9,158 5,251 4,519 4,305	12.93 4.47 2.56 2.21 2.10	attective predicates a singe + breakthrough single creative breakthrough creati invenient	2.400 123 54 122	5,41 5,29	categoria de la composition nuencion breakthrough wi discovery breakthrough wi	Lan 1.433 2 20 as the invention 21 as the discover	0.70 6.35 6.21 m 6.13	Dessettors of D science ndp beetles band the bands br tomorrow	reakthrough 2,820 12 12 12 12 12 12 12 12 12 12 12 12 12	1.38 6.11 6.02 5.35 4.93	discovery St. Cynengy TK is a bre technology + finding			21 12 122 19	6.40 6.20 6.10 5.47
Descolitional phrases Descolitional phrases Descolitional phrases Descolitional? 	74,948 25,477 9,158 5,251 4,519 4,326 3,312	12.93 4.47 2.56 2.21 2.10 1.62	Attective predicates of singe + breakthrough single creative breakthrough creati instinent breakthrough was in	C Dereakthe 2.400 153 54 Net 29 millionit 11	5,41 5,29	corporation incorporation invention breakthrough wi biocovery breakthrough wi nilestone	Lam LAJJ 2 20 as the inventio 21 as the discove 22	0.70 6.85 6.21 m 6.13 ry of 4.50 4.59	Dessessors of science ndp beetles band the band's br tomerrow bomerrow	reakthrough 2,820 12 12 12 12 12 12 12 12 12 12 12 12 12	1.38 6.11 6.02 5.35 4.93	discovery B. Cynergy TK Is a bre technology + finding generator			21 12 122 19	6.10 5.47 5.45
Desitional shoes Treasthough W of Treasthough is Treasthough to Treasthough Treasthough G Treasthough G Scientificath G as Treasthough	74,968 25,477 9,155 5,251 4,519 4,205 3,312 2,083 1,810	12.93 4.47 2.56 2.21 2.10 1.62 1.02	Attentive predicates o singe + breatthrough single cettie breatthrough creat breatthrough vas is forthcoming unitedy	1 Dreakth 2,400 123 54 twe 22 minerit 11 32	5.41 5.45 5.45 5.46 5.46	categoriation breakthrough vi incorporation invention breakthrough wi filestone sevelopment	Lam LAJJ 2 20 as the inventio 21 as the discove 22	0.70 6.85 6.21 m 6.13 ry of 4.50 4.59	breaktbrouch science ndp beaktes band the bands br bomorow's b vright	reskthrough 2,820 12 20 20 eskthrough 52 reskthrough	1.38 6.11 6.02 5.35 4.93 4.69 4.36	Association of the second of t	sktroug		21 122 19 14 9 29	6.40 6.20 5.47 5.45 5.38 5.23
repositional phrases Descriptional phrases and "breakthrough" and "breakthrough" and "breakthrough" Descriptional" of an and the phrase through " Descriptional" of an Descriptional" of an and the phrase through " Descriptional" of an and the phrase through the phrase the phrase through the phrase through the phrase throu	74.968 25.477 9.155 5.251 4.519 4.325 3.212 2.053 1.810 1.659	12.93 4.47 2.56 2.21 2.10 1.62 1.02 0.88 0.81	Affective predicates o singe + breakthrough single creative breakthrough creati inminent breakthrough ves in forthcoming	1 Denakthr 2.400 123 54 ket 22 minient 11 32 kety .	5.41 5.41 5.41 5.44 3.44	cale breakthrough" is a incorporation invention breakthrough wi blockery breakthrough wi intertone severopment breakthrough wi	1.433 2 20 as the inentia 21 as the discove 2 2 as the develop 14	0.70 6.85 6.21 m 6.13 ry of 4.60 4.59 ment of	Desicitivo units accence ndp Destiles Dend the band's br tomorrow Lomorrow's b wright wright's brea	reskthrough 2,820 12 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	1.38 6.11 6.02 5.35 4.93 4.69 4.36	us is a Decekthrough discovery B: Cynengy TK is a bre- technolog + finding generator cambogia research	sktroug		21 122 19 14 9 29	6.40 6.20 5.47 5.45 5.38 5.23
accessional phone accessional phone accessional phone of Density from to Density form to Density form to Density form the Density of the test phone for the Density of the	74.968 25.477 9.155 5.251 4.519 4.305 3.212 2.083 1.810 1.659 1.544	12.93 4.47 2.56 2.21 2.10 1.62 1.02 0.88 0.81 0.75	Affective predicates o singe + breakthrough single creative breakthrough creati inverser fortscoming unitely breakthrough is unit	4 Dereakth 2.400 122 54 ket 29 milliont 11 32 kety. 142	5.41 5.41 5.41 5.41 5.29 3.46 2.70	cataneous breakthrough is a morporation invention breakthrough w breakthrough w miletone Sevelopert breakthrough w ntreduction	L433 2 20 as the Invention 21 as the discover 22 as the develop 14 25	0.70 6.85 6.21 m 6.13 ry of 4.60 4.59 ment of 4.21 3.30	Desicitivo units accence ndp Destiles Dend the band's br tomorrow Lomorrow's b wright wright's brea	reskthrough 2,820 12 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	1.38 6.11 6.02 5.35 4.93 4.69 4.36 ms you	Alening discovery B. Cynengy TK is a bre- technology + finding generator cambogia research vision care and rese	sktroug		21 12 12 19 14 29 najor breakthrough	8,40 6,20 6,10 5,47 5,45 5,38 5,23 5,23 5,23 5,23 4,91
Propositional phones and "breakthrough" In of "breakthrough" In to "breakthrough" In the Breakthrough" Breakthrough" of In an Breakthrough" of In Breakthrough" of In Seekthrough" of In In the Breakthrough" of In In the Breakthrough of In In the	74.968 25.477 9.155 5.251 4.519 4.525 3.212 2.051 1.610 1.659 1.544 1.253	12.93 4.47 2.56 2.21 2.10 1.62 1.02 0.88 0.81 0.75	Attestive predicates of singe + breatthrough single creative breatthrough oreat imminent breatthrough vesi in forthcomine unitely breatthrough is unit possible +	1 Denakthr 2.400 123 54 ket 22 minient 11 32 kety .	5.41 5.29 5.41 5.45 3.45 2.70	citometri incorporation meention breakthrough wi sectory breakthrough wi nitestane development breakthrough wi development breakthrough wi nitestane development breakthrough wi development breakthrough wi developme	Len 1.433 20 as the invention 21 as the discover 22 as the develop 14 25 22 22 23 24 25 25 20 25 25 25 25 25 25 25 25 25 25	0.70 6.85 6.21 m 6.13 ry of 4.60 4.59 ment of 4.21 3.30	breakthrough science ndp beates band the bands br bomorow's b wright minister minister	reakthrough 2,420 10 50 50 50 50 50 50 40 reakthrough 12 4through unle 8 2	1.38 6.11 6.02 5.35 4.93 4.69 4.36 mis you 3.86	us is a December discovery b: Cynengy TK is a bre- technology finding genesitor cambogia research vision care and rese innovation	sktroug		21 12 12 19 14 29 najor breakthrough	6.40 6.20 5.47 5.45 5.38 5.23 , the program
Descriptional phrases Descriptional phrases of Terestimousti us Descriptional Sectory of the Descriptional of the	74.968 25.477 9.155 5.251 4.519 4.305 3.212 2.083 1.819 1.659 1.544 1.283 1.319	12.93 4.47 2.56 2.21 2.10 1.62 1.02 0.88 0.81 0.75 0.68	Adjustive predicates of single + breakthrough single cretitive breakthrough vess is forthcoming unlikely breakthrough is unli posible + 3d	22 000 122 122 123 123 123 123 123 124 121 122 122 122 122 123 123 123	5.41 5.41 5.29 5.45 3.46 2.70 2.54	citamina neorporation meetion breakthrough wi discovery breakthrough wi mitestane sevelopment breakthrough wi mitestane breakthrough wi mitestane	Lean 1.423 20 as the invention 21 as the discover 22 as the develop 14 25 20 the use of	0.70 6.85 6.21 m 6.13 ry of 4.60 4.59 ment of 4.21 3.30	beekthrough scence ndp beektes band the bands br bomorrow's b wright wright dewin seciety	reakthrough 2,820 12 29 20 eakthrough 12 athroughs unli 8 21 21	1.38 6.11 6.02 5.35 4.03 4.69 4.36 mis you 3.86 3.81	sublecting discovery Cynergy TK is a breat technology + finding generator cambogia research vision care and rese innovation arrest	sktroug		21 12 12 19 14 29 najor breakthrough	8,40 6,20 5,47 5,45 5,38 5,23 , the program 4,91 4,91
Dependitional shores Terminoutin' Simi in of Terminoutin' Simi in to Terminitrough the Terminitrough Terminoutin' of m must Terminitrough Terminoutin' of m must Terminitrough Terminoutin' of m must Terminitrough Terminit	74,968 25,477 9,158 5,251 4,519 4,529 1,525 1,810 1,659 1,545 1,545 1,545 1,545 1,545 1,545	12.93 4.47 2.56 2.21 2.10 1.62 1.02 0.88 0.81 0.75 0.68 0.64	Affective predicates of singe + breathrough single creative breathrough was is forthcoming unitely breathrough is uni- possible + 3d breathrough 35	4 Denskthr 2.400 122 54 here 22 22 122 123 123 124 124 124 124 124 124 124 124	5.41 5.41 5.29 5.45 3.46 2.70 2.54	citizente meruperation nivertion breakthrough w breakthrough w breakthrough w breakthrough w breakthrough w breakthrough is breakthrough is	Longits Long 1,432 20 as the invertisi as the discover 22 as the develop 23 as the develop 20 20 20 20 20 20 20 20 20 20	0.70 0.85 6.21 m 6.13 ry of 4.60 4.59 ment of 4.21 3.30 3.25	beekthrough science ndp beektes band the bands br bomorrow tomorrow tomorrow tomorrow tomorrow tomorrow tomorrow tomorrow tomorrow tomorrow tomorrow	reaktrough 12 12 10 10 10 10 10 10 10 10 10 10 10 10 10	1.38 6.11 6.02 5.35 4.03 4.69 4.36 mis you 3.86 3.81	sources sources Cynery BL Cynery TK is a breact technology + finding generator cambogia research vision care and rese finding series time sectors time sectors time sectors time sectors time time sectors time time sectors time time time time time time time time time time time time time time time	sktroug		21 12 12 19 14 29 najor breakthrough	6.40 6.20 5.47 5.45 5.38 5.23 , the program 4.91 4.91 4.91
Descriptional phrases Descriptional phrases Descriptional phrases and Descriptional Descriptional Descriptional Descriptional Descriptional Descriptional Descriptional Descriptional Descriptional Descriptional Descriptional Descriptional Descriptional Descriptional	24.968 25.477 9.155 5.251 4.519 4.519 4.519 1.212 2.053 1.810 1.659 1.544 1.253 1.319 1.253 1.319 1.259 1.001	12.93 4.47 2.56 2.21 2.10 1.62 1.02 0.88 0.81 0.75 0.68 0.64 0.63	Affective predicates of singe + breathrough single creative breathrough creati instinent breathrough vas is fortbooming unkliely breathrough is unit possible + 3d breathrough 3D reatmit breathrough notur	1 Denskthr 2.400 122 54 her 22 meninerst 11 32 herty. 142 12 12 11	South: 1 1.17 6.22 5.41 6 5.29 6 3.45 3 2.70 2.54 2.44 7	citizensen breekthrouch' is a ncorporation meerdion breakthrough wi biotowery breakthrough wi breakthrough is exit breakthrough is exit breakthrough is exit	Longits Long 1,432 20 as the invertisi as the discover 22 as the develop 23 as the develop 20 20 20 20 20 20 20 20 20 20	0.70 0.85 6.21 m 6.13 ry of 4.60 4.59 ment of 4.21 3.30 3.25	bookthrough science ndp bend bend the bands br bomorow: b wright mingits bes einstein devrin seciety Society's line group	reakthrough 12 20 20 20 20 20 20 20 20 20 20 20 20 20	1.38 6.11 6.02 5.35 4.93 4.69 4.36 5.35 3.85 3.81 3.48	Usening discovery b, Cynergy TK is a breaching generator cambogia cambogia cambogia cambogia research vision care and rese innovation arvest tes estract announcement	sktroug		21 13 19 19 19 20 20 20 20 20 20 20 20 20 20 20 20 20	5.40 6.20 5.47 5.45 5.38 5.23 , the program 4.91 4.91 4.92 4.92 4.92 4.92
Constitute and Constitu	74.968 25.477 9.158 5.251 4.519 4.208 3.312 2.083 1.810 1.659 1.544 1.383 1.219 1.219 1.219 1.200 1.200 1.200	12.93 4.47 2.56 2.21 2.10 1.62 1.02 0.88 0.81 0.75 0.68 0.64 0.43 0.49 0.45	Adjustive predicates a singe + breathrough dreat inminet breathrough rest is forthcoming unitely breathrough rest breathrough notin top	4 Denskthr 2.400 122 54 here 22 22 122 123 123 124 124 124 124 124 124 124 124	S.47 S.47 5.41 5 5.29 1 5.29 1 3.45 1 2.70 2 2.54 1 2.44 1 2.30 1	citizensen incorporation meertion breakthrough wi discovery breakthrough wi mitestone breakthrough wi werduntten breakthrough is werhoology is werhoology is	Longits Long 1,432 20 as the invertisi as the discover 22 as the develop 23 as the develop 20 20 20 20 20 20 20 20 20 20	0.70 0.85 6.21 1 6.13 ry of 4.60 4.59 ment of 4.21 3.30 3.28 3.26 3.00	Desettive units accence indo Destinas Dente Dente Denterrow Lonerrows to wright wrights breat einstein derwin society's Breat group the group's b	reakthrough 2.420 12 10 20 eakthrough 41 41 eakthrough 1 22 akthrough 1 22 reakthrough 1 22 reakthrough 1	1.38 6.11 6.02 5.35 4.93 4.69 4.36 3.81 3.48 2.37	In the second se	sichroug earch is a	n and the r	21 12 19 14 29 20 20 20 20 20 20 20 20 20 20 20 20 20	6.40 6.20 5.47 5.45 5.38 5.38 5.38 5.38 5.38 5.38 5.38 5.3
Descriptional phrases Descriptional phrases Descriptional phrases Descriptional for the setting of the setting the setting of the setting the setting of the setting of the Descriptional for the Descriptional for the Descriptional for the setting of the Descriptional for Descriptional for Descri	74.968 29.477 9.153 5.251 4.519 4.205 3.312 2.083 1.810 1.659 1.559 1.233 1.219 1.233 1.219 1.233 1.219 1.233 1.219 1.233 1.219 1.233 1.219 1.233 1.219	12.93 4.47 2.56 2.21 1.62 1.62 0.68 0.61 0.75 0.64 0.49 0.45 0.45 0.45	Adjustive predicates of snge + breakthrough single creative breakthrough creati invinent breakthrough vas is forthcoming untilety breakthrough 10 reakthrough notur top breakthrough top	1 Denskthr 2.400 122 54 her 22 meninerst 11 32 herty. 142 12 12 11	Counth: 1.97 0.22 1 5.41 5.29 3.95 6 3.95 6 2.56 0 2.56 0 2.56 0 2.56 0 2.56 0 1.00 0 1.0	corporation meorporation meansubhrough wi biscovery breakthrough wi heistone sevelopment breakthrough wi heisduction belity breakthrough is equal breakthrough is echology meus	Lange Courses 20 20 20 20 20 20 20 20 20 20	0.70 6.85 6.21 71 6.13 72 of 4.60 4.59 metric of 4.21 3.30 3.28 3.26 3.00 2.52	benefitive units science ndp bendies bend the band's br bomorrow's b wright tomorrow's br wrights brea einstein devvin seciety Society's Bre group the group's b perty	reakthrough 2.420 12 20 12 12 12 12 12 12 12 12 12 12	1.38 6.11 6.02 5.35 4.93 4.69 4.36 5.35 3.85 3.81 3.48	Alerning discovery B, Cynergy TK is a breach technology + finding generator cambogia research vision care and rese innovation arrest tes estract announcement vision . Pfber Viagra is th	sichroug earch is a	n and the r	21 12 12 19 9 29 29 29 29 8 8 8 8 2 2 10 10 10 10 10 10 10 10 10 10 10 10 10	6.40 6.20 6.10 5.47 5.45 5.23 , the program 4.91 4.91 4.91 4.92 4.92 4.93 4.93 4.93 4.93 4.93 4.93 4.94
Constitutional advances Constitutional advances Constitutional Constitut	74.968 29.477 9.158 5.251 4.519 4.529 3.212 2.083 1.810 1.659 1.545 1.2383 1.219 1.298 1.001 9.298 1.001 9.247 5.77	12.93 4.47 2.56 2.210 1.62 1.62 0.88 0.81 0.75 0.68 0.64 0.49 0.49 0.49 0.49 0.36	Adjustice predicates of single + breathrough single creatily breathrough read breathrough vals is forthcoming untilely breathrough to unit possible + 3d breathrough nature top breathrough nature top	22 122 122 123 124 124 124 124 124 124 124 124	2.300 t 1.97 0.222 1 5.41 5.29 0.45	country of the second s	Longits Long 1,432 20 as the invertisi as the discover 22 as the develop 23 as the develop 20 20 20 20 20 20 20 20 20 20	0.70 6.85 6.21 6.13 77 of 4.50 4.59 4.21 3.28 3.28 3.28 3.28 3.28 3.25 2.52 2.15	beekthrough Connect ndp beektes band the bands br bomerrow	reakt/booker 2,420 12 12 12 12 20 eakt/broughs 12 20 eakt/broughs 21 22 akt/roughs unle 2 2 akt/roughs 22 akt/rough 22 akt/rough 22 22 akt/rough 22 22 akt/rough 22 22 23 24 24 24 24 24 24 24 24 24 24 24 24 24	1.38 6.11 6.02 5.35 4.93 4.69 4.36 3.81 3.48 2.37 2.25	teshening teshening discovery bi Crempy TK is a brea technology + finding generator cambogia research vision care and rese innovation arrest teshenita extract annoucement visars . Fiber Viagra is th device	sichroug earch is a	n and the r	21 12 12 19 9 9 9 8 8 8 8 8 9 14 9 9 8 8 8 8 9 14 19 9 8 8 8 8 8 9 14 15 14 15 14 14 15 14 15 14 15 14 15 14 15 14 15 14 15 15 15 15 15 15 15 15 15 15 15 15 15	6.40 6.20 8.10 5.47 5.45 5.38 5.23 , the program 4.91 4.91 4.91 4.91 4.92 4.92 4.92 4.92 4.92 4.92 4.92 4.92
Descriptional shares Sector Should shares Sector Should shares Sector Should shares Sector Should share Sector Share	24.968 25.477 9.158 5.251 4.519 4.309 4.308 1.212 2.083 1.212 1.649 1.544 1.383 1.219 1.200 1.544 1.383 1.210 1.2000 1.2000 1.2000 1.20000000000	12.93 4.47 2.56 2.21 1.62 1.62 1.62 0.88 0.84 0.43 0.49 0.45 0.44 0.45 0.44 0.45 0.44 0.45 0.44 0.45 0.44 0.45 0.44 0.45 0.44 0.45 0.44 0.45 0.44 0.45 0.44 0.45 0.44 0.45 0.45	Adjustive predicates a singe + breakthrough single creative breakthrough creati inminent breakthrough reas is forthorowing unitary breakthrough top reature breakthrough return top breakthrough top reav breakthrough top reav breakthrough top	(1) Dereskthe 2,400 132 	00000000000000000000000000000000000000	citizenter trendstaturent in memperation mention breakthrough with society breakthrough with breakthrough is result breakthrough is resu	Lange Courses 20 20 20 20 20 20 20 20 20 20	0.70 6.85 6.21 m 6.13 ry of 4.60 4.59 ment. of 4.21 3.30 3.28 3.26 3.00 2.515 1.45	beekthrough science ndp beekles bend the bands b bend the bands b bend benerrow bone	reaktroach 2,820 12 10 10 10 10 10 10 10 10 10 10 10 10 10	1.38 6.11 6.02 5.35 4.03 4.56 4.36 4.36 8.81 3.48 2.37 2.25 2.18	sub-ening discovery Cynengy TK is a bre- technology + finding generator cambogia research vision care and rese timovation arrest tas extract anouncement vision cambogia research vision care and rese timovation arrest tas extract anouncement vision agreement	sichroug earch is a	n and the r	21 12 12 14 29 14 29 29 29 20 20 20 20 20 20 20 21 21 21 21 21 21 22 22 22 22 23 24 23 24 23 24 23 24 23 24 24 24 24 25 24 24 24 25 24 24 24 25 24 25 24 25 24 25 26 26 26 26 26 26 26 26 26 26 26 26 26	6.40 6.20 6.10 5.45 5.38 5.23 5.23 5.23 5.23 4.91 4.91 4.91 4.91 4.91 4.92 4.95 4.44 4.29 4.29 4.37
Descriptional phone Treatingation T	24.968 25.477 5.158 5.251 4.519 4.519 4.519 4.2051 1.810 1.654 1.253 1.2	12.93 4.47 2.56 2.21 1.62 1.62 0.88 0.81 0.49 0.49 0.49 0.49 0.49 0.49 0.49 0.49	Affective predicates of single + breathrough single creative breathrough creati inninest breathrough vas is breathrough to unit possible + 3d breathrough top rear top breathrough top rear innine creating rate	2.400 123 54 bet 22 ministerit 11 12 12 12 13 13 13 13 13	2.70 5 1.97 5 5.29 5 3.45 5 2.99 5 2.54 5 2.99 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	columnia constitutions of the second meetion breakthrough with instation breakthrough with breakthrough with breakthrough is breakthrough is breakthrough is schoology service top page settor breakthrough is breakthrough is schoology service top settor breakthrough is breakthrough is schoology service top settor breakthrough is breakthrough is breakthrough is schoology service top settor breakthrough is breakthrough is brea	the state of the result of the	0.70 6.55 6.21 6.13 6.13 77 of 4.60 4.59 ment of 4.21 3.00 3.28 3.26 3.00 2.512 2.155 1.45 1.22	bookthrough science indp bend the bands br bomorrow bomorow bomoro	reakthrough 2,220 12 12 12 12 12 12 12 12 12 12 12 12 12	1.38 6.11 6.02 4.93 4.94 4.36 4.93 4.69 4.36 5.35 5.35 5.35 5.35 5.35 5.35 5.35 5	Alening discovery bit Cynergy TK is a breat technology + finding generator cambogia research vision care and rese innovation arrest tas entract announcement viagra _Proce Viagra is the device agreement procedure	akthroug sarch is a	hrough c	21 12 12 19 9 9 9 8 8 8 8 8 9 14 9 9 8 8 8 8 9 14 19 9 8 8 8 8 8 9 14 15 14 15 14 14 15 14 15 14 15 14 15 14 15 14 15 14 15 15 15 15 15 15 15 15 15 15 15 15 15	6.40 6.20 8.10 5.47 5.45 5.38 5.23 , the program 4.91 4.91 4.91 4.91 4.92 4.92 4.92 4.92 4.92 4.92 4.92 4.92
resolutional shraves resolutional shraves of Pressimouth' form of Pressimouth' of man of Pressimouth' of man of Pressimouth' of man of Pressimouth' of man and Pressimouth' of man of Pressimouth' of man or Restance of the man of the man	24.968 25.477 9.158 5.251 4.519 4.309 4.308 1.212 2.083 1.212 1.649 1.544 1.383 1.219 1.200 1.544 1.383 1.210 1.2000 1.2000 1.2000 1.20000000000	12.93 4.47 2.56 2.21 1.62 1.62 0.88 0.68 0.68 0.68 0.65 0.68 0.45 0.25 0.25	Affective predicates of single + breathrough single creative breathrough creati inninest breathrough vas is breathrough to unit possible + 3d breathrough top rear top breathrough top rear innine creating rate	2.400 123 54 bet 22 ministerit 11 12 12 12 13 13 13 13 13	0004011 2 1.172 0.22 1 5.41 5.29 3.45 4 2.70 4 2.70 1 2.54 1 2.70 1 1.90 1 1.97 6 1.93 1 1.90 1 1.97 6 1.93 1 1.93 1 1.93 1 1.93 1 1.94 1 1.94 1 1.95 1	citizenter breakthrough is breakthrough w breakthrough w breakthrough w htroduction breakthrough is evel breakthrough is evel b	Lange Courses 20 20 20 20 20 20 20 20 20 20	0.70 6.85 6.21 m 6.13 ry of 4.60 4.59 ment. of 4.21 3.30 3.28 3.26 3.00 2.515 1.45	Desistance of 2 Science ndp Desities Dead the bands br bomerrow bomerow bomerrow bom	reaktroach 2,820 12 10 10 10 10 10 10 10 10 10 10 10 10 10	1.38 0.11 1.38 0.12 1.48 0.12 1.48 0.18 0.19 1.48 0.18 1.48 0.18 1.48 0.18 1.48 0.18 1.48 0.18 1.48 0.18 1.48 0.18 1.48 0.19 1.48 0.19 1.48 0.19 1.48 0.10 1.48 0.10 1.48 0.10 1.48 0.10 1.48 0.10	sub-ening discovery Cynengy TK is a bre- technology + finding generator cambogia research vision care and rese tinovation arrest tas exitant announcement vision agreement signeement	akthroug sarch is a	hrough c	21 12 12 14 29 14 29 29 29 20 20 20 20 20 20 20 21 21 21 21 21 21 22 22 22 22 23 24 23 24 23 24 23 24 23 24 24 24 24 25 24 24 24 25 24 24 24 25 24 25 24 25 24 25 26 26 26 26 26 26 26 26 26 26 26 26 26	6.40 6.20 6.47 5.45 5.35 5.23 , the program 4.91 4.92 4.95 4.49 4.49 4.49 4.49 4.49 4.49 4.49

Figure 1: An example of a Word Sketch

Remember, if you have any ideas and/or suggestions regarding helpful resources or tools you would like to see featured, please e-mail me at jzetzsche@internationalwriters.com.

Word Sketches: This is where the program got its name, and it's what Kilgarriff brought to the table. A word sketch is a summary of a word's grammatical and collocational behavior (collocational refers to the analysis of how often a word co-occurs with other words or phrases. (See Figure 1 on page 36.) Since the data in the corpora is lemmatized (i.e., words are analyzed so they can be brought back to their base or dictionary form), the results are a lot more meaningful than what most of our translation environment tools provide when they're unable to relate different forms of one word to each other. Another word sketch option that Sketch Engine offers is the comparison of word sketches of similar words.

Thesaurus: The ability to retrieve a detailed list or a graphical word cloud with similar words, including links to create reports on word sketch differences for those terms to understand the exact differences in actual usage.

Concordance: Searches for single words, terms, or even longer phrases. Since the data in the supported languages is tagged, it's also possible to search for specific classes of words or specific classes of words that surround the word in question.

Parallel Corpus: Retrieval of bilingual or multilingual sets of words or phrases within the contexts. (See Figure 2.) Presently this is available only for onscreen data viewing, but it will soon be offered as downloadable data. This is especially helpful when uploading your own translation memories (see below).

Word Lists: The possibility of creating lists of words and the number of occurrences, either as lemmas (the base form of each word) or in each word form.

Creating Your Own Corpus: This is likely the most exciting feature for translators. You can either upload your own translation memories or use the tool's own search engine mechanism (which relies on Microsoft Bing) to create a list of bilingual websites that contain the terms that are relevant to your field, have them automatically align, and form a corpus. I don't need to explain to you the possibilities this offers to translators who don't have Query convoluted 40 > Priter by aligned corpus 33 > Priter by aligned corpus 32 > Priter by aligned corpus 4 (0.07 per relition) (0

design-dif-	EUROPARZ, Endtah	Marker 47.	EUROPAR, 7. German	and.	EUROPART, halas	Adm dt.	Padei zevan de
	the inertier black? I support that, in some convertient way, those who want was interference from		In table dar Landsnitt, dass in Typerschlut antergrenzen tellen distangen, die songer Unterstandig distangen, die songer Unterstandig Robert des Jahren der State anter die Jahren anderstande der State anter Antergehöhten der State anter die serbauten und Rechtwarte und anter des State anter die State antergehöhten des State anter-		Superto che, il un ando sicurito controlito cuello sicurito controlito cuello restarte esta cuella cuello interferensa finale regitta della portuciono di interferensa finale regitta della cuella cuella della cuella cuella discontro il cuella cuella esta cuella cuella cuella anteresa per socialitta ei della cuella cuella cuella anteresa anteresa attana.		- parties generative standard Co. 1972 doorses in registeric foreign and the standard management foreign and the standard and the standard and the standard domestic re- solution and the standard particular and standard participation and standard participation a
Arming-08	ensities consume parameter, but excessively conversed nucl-decade $\sim_{2}\!$	Jahring-OB	Auth der Infelliewerb und die eligenetie Auflahrung der Verbraucher denem dem Verbraucherschutz Oberwongsteht hatt, dem Verbraucherschutz Hört, «Jo-	309-68	Concontences ed inducatione generationals det consumeration anne anche utilit per occrescenne în batola, ere norme accescimente complicate henno (* entreto complicate henno (* entret	/j84g-68	Resturbed a light a velocitie a optima velocitie (a becay tentor) venocriti optivore annariam fa, s.e. redminier (a solite coastro (fa, cito)

Figure 2: The parallel corpus feature

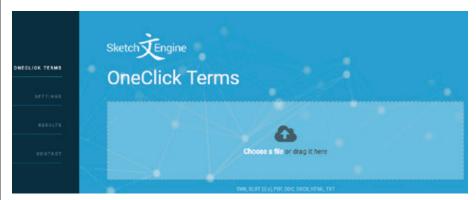


Figure 3: Preview of OneClick Terms

the privilege of having high-quality translation memories or termbases for a particular subject matter that they need to translate. As a logical extension of this feature, not only can you perform any of the functions mentioned earlier, but it's also possible to run a keyword search on the user-created corpus, identify the terms that are relevant, and download that into an Excel or TBX file. This feature is currently available for Chinese, Czech, Dutch, English, French, German, Italian, Japanese, Korean, Polish, Portuguese, Russian, and Spanish. The bilingual version of this is just around the corner.

By the way, you can find an example of the up-and-coming increased userfriendliness of Sketch Engine in OneClick Terms (terms.sketchengine.co.uk) that allows you to extract terms from TMX, XLIFF, PDF, DOC, DOCX, HTML, or TXT files in essentially one or two clicks. (See Figure 3.)

Translators have been one of the primary target groups for the makers of Sketch Engine. One immediate result of that focus is the availability of a plug-in for SDL Trados Studio (see http://bit.ly/SDL- appstore-sketch and http://bit.ly/ user-guide-sketch). The plug-in itself is free, but it requires a trial or paid registration to be usable. It allows you to perform collocation, thesaurus, and concordance searches and will soon offer term extraction. According to Ondřej Matuska at Sketch Engine, talks with makers of other translation environment tools are under way to offer plug-ins or add-ons for those tools as well.

Can you believe you've never heard about this tool before? Well, maybe you were quicker than I to find this, but the good thing is that now we all know. •



Jost Zetzsche is chair of ATA's Translation and Interpreting Resources Committee. He writes the "Geekspeak" column for *The ATA Chronicle*. He is also the coauthor of *Found in Translation*:

How Language Shapes Our Lives and Transforms the World, a robust source for replenishing your arsenal of information about how human translation and machine translation each play an important part in the broader world of translation. Contact: jzetzsche@internationalwriters.com.