Scientific Report of Short Term Scientific Mission

COST STSM Reference Number: COST-STSM-IS1305- 24203

Period: 25-01-2015 to 07-03-2015

Duration: 30 working days

COST Action: IS1305

STSM type: Regular (from Slovenia to Poland)

STSM Title: Automatic Extraction of Polish Multi-Word Expressions

Guest/STSM applicant: Cyprian Laskowski, University of Ljubljana

Host: Prof. Piotr Żmigrodzki, Instytut Języka Polskiego, piotr@ijp-pan.krakow.pl

1. Purpose of the STSM

The main goal of my STSM was to take methods used in Slovenia for automatically extracting multi-word expressions (MWEs), apply them to Polish and demonstrate the usefulness of this approach at the Institute of the Polish Language (IPL) in Kraków, where the Great Dictionary of the Polish Language (GDPL) is compiled and edited. A secondary and more ambitious goal was to visit the Linguistic Engineering Group (LEG) in Warsaw and the plWordNet group (PWNG) in Wrocław and apply automatic methods for assessing the degree of semantic compositionality to the multi-word expressions I had extracted.

2. Description of the work carried out during the STSM

I started and ended my trip at the IPL. On the first day, I was warmly welcomed by Prof. Żmigrodzki, introduced my project goals and Sketch Engine to the GDPL staff, and had several discussions with several staff members to understand their current tools, methods and perspectives for handling MWEs. Then I set up my Sketch Engine account for the work, selecting a testing corpus already available on Sketch Engine which was of an appropriate size and which used the same tagset as the National Corpus of Polish (NCP). I then engaged the bulk of the work, preparing the Polish sketch grammar.

I started with the well-established Slovene sketch grammar, which was developed by Dr. Krek and colleagues in Slovenia a few years ago, and transformed the Slovene sketch grammar for Polish. This involved mainly 3 parts. First, I changed the regular expressions for matching elements in the tagged corpus to fit the tagset used my NCP (e.g., [tag="(subst|depr|ger):.*"] for collocate nouns). Second, I adjusted the gramrels to reflect the differences between Polish and Slovene grammar (e.g., adjectives come before nouns in Slovene, but can come before or after nouns in Polish). Third, I systematically renamed the macros and gramrels (grammatical relations) for Polish (e.g., rzecz_d_1" for head nouns in genitive (dopełniacz) case). I iteratively made changes, uploaded and recompiled the changed sketch grammar, and tested the results on a few words in the Sketch Engine interface. I also adjusted the Python script provided by Sketch Engine for batch extraction and the related configuration files for Polish and used the script to extract the collocations for a few words in XML format. I had a few conceptual and technical issues along the way, but I was in touch with my Slovene colleagues about pecularities of the Slovene sketch grammar and the Sketch Engine support staff about technical difficulties.

Once I had more or less prepared the Polish sketch grammar, I presented the interim results to the GDPL staff, showing them a few gramrels along with the tabulated results in the Sketch Engine interface for a few words. I got some useful feedback concerning a few gramrels, and was asked if I could also adjust the organisation of the gramrels to that of their editorial system. This effectively meant restructuring and renaming my gramrels to be in a 1-to-1 relationship with their collocation categories. As they had given me access to their dictionary editing pages, I then studied these categories more closely and reorganised my gramrels accordingly. To that end, it was very useful that the desk and space they provided me with was in the same room as two of their dictionary editors, who clarified various details and problematic cases for me during my stay. Some of these turned out to be rather subtle. For instance, gerunds are generally treated as verbs, so that the collocates of gerunds were listed under the corresponding verb entry, but gerund forms were included among the noun collocates of adjectives. This meant that the final gramrels were not as simple and symmetric as the initial versions.

At this point, I went to Warsaw and Wrocław to address the other part of my STSM, which was to attempt to automatically detect the semantic compositionality of collocations. When I came to Warsaw, I met with Prof. Przepiórkowski, with whom I had arranged to come and work on this problem during this period. Unfortunately, as he had warned me, this was a very busy period for him, and in fact even busier than he had expected. Also, it turned out when we talked that although his group was interested in this problem, they had not yet developed tools or methods to address it. Therefore, although I presented my STSM goals to his group and also attended an interesting presentation on the valience dictionary they have been developing, I quickly switched my focus to Wrocław. I contacted Prof. Piasecki's team and asked them about their approach to the problem. They were very helpful and communicative, so I spent much of my week in Warsaw familiarising myself with their methods remotely.

Then I went to Wrocław and communicated more closely with Prof. Piasecki's enthusiastic PWNG team. The basis of their approach is to statistically compare the distribution of words with the distributions of MWEs including those words. They look not only at text windows, but at grammatical relations that the words or MWEs enter into with specific other words. The basic idea is that if a MWE is semantically compositional, then the distribution of the head word is generally more likely to be similar to the distribution of the entire MWE. They also used additional heuristics in combination with this method, such as whether the word order of a MWE is fixed and whether it can have other words inserted in between. Their studies had shown that the method they were developing for noun-adjective collocations was already achieving good results. I therefore wanted to try to apply their tools to my data to get semantic compositionality values for a small sample of my noun-adjective MWEs.

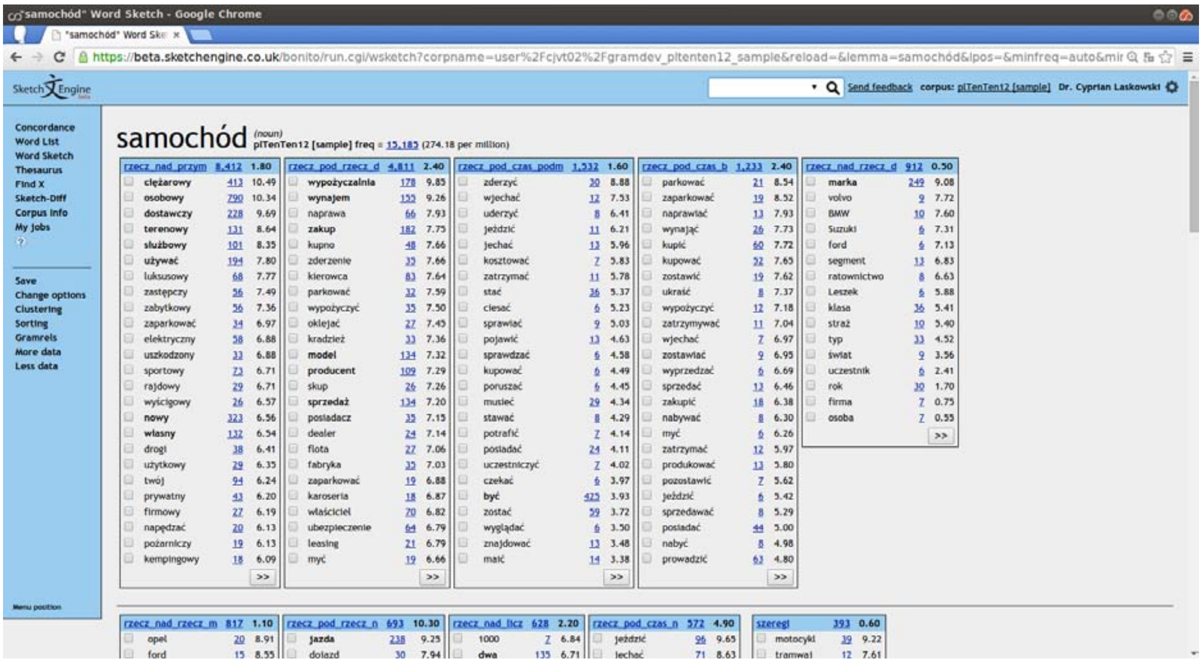
Unfortunately, this turned out to be too ambitious a task to do in such a short period of time. First of all, their method relies on having data for several grammatical relations for both the headwords and the collocations on a large corpus. As I had only been using a small corpus, I had to make a simplified version of my sketch grammar and fetch the data on a large corpus. I got permission from the Sketch Engine staff to use their largest corpus, but extracting these results took several days. The PWNG did offer me to use their own corpus data, but I felt I should use the data I had gathered; in hindsight, this was probably a mistake. Second, I needed to convert the data from Sketch Engine to a suitable format for the PWNG's SuperMatrix tool, but as I had not done this before, I made a subtle but significant mistake in the conversion, which greatly reduced the validity of my data. Unfortunately, by the time Prof. Piasecki noticed and informed me of my mistake, the data had already been mostly processed by SuperMatrix and there wasn't time to do it again. Third, in my attempt to select interesting nouns and collocations to test, I had selected mostly head nouns which were polysemic, which reduces the reliability of their method. Fourth, again due to time limitations, I wasn't able to incorporate the other heuristics mentioned above(word order, separability) into the method. Fifth, as I had only wanted to make a small demonstration and was short on time, I had only selected a small number of collocations to test, but this did not allow for a proper statistical analysis of the results, which was particularly problematic given the other compromises and limitations listed above. As a result, my results were not adequate and I have to defer this ambitious goal to the future. My experience in Wrocław taught me that this problem can be effectively tackled in clever ways, and that the PWNG team is wonderful to work with, but that this is indeed a difficult problem and cannot be solved without the proper considerations and time.

After my stay in Wrocław, I returned to Kraków. I did a few more tests and improvements of my Polish sketch grammar. Then I thought about how I could most effectively show them its potential usefulness, and decided to do so directly in the context of their internal website. I mapped the XML collocations downloaded for a sample word to JavaScript format, and added them to a saved copy of their editor page for a particular word. I also added an HTML button on this page, which, when clicked, showed the list of downloaded MWEs at the top of the window, along with checkboxes and buttons, which would allow them to select which meanings of the word the MWE should be assigned to and buttons for pasting the word directly in the appropriate MWE textarea (based on the 1-to-1 mapping between gramrel and collocation category). I implemented this partially for one example, enough to show how automatically fetched collocations could be usefully integrated into their existing internal editor webpages.

Finally, I gave a concluding presentation and demo. I showed the final sketch grammar, along with a few examples, and one extended example showing how my sketch grammar found most of the collocations they had listed in their dictionary for a common word, even with my small testing corpus. Then I showed them my edited version of their editor page. The response was enthusiastic, especially once they saw their edited webpage and saw concretely how such methods could help them directly with their daily work. Prof. Żmigrodzki said that I have given them a lot to think about.

In all, I think the main goal of my STSM was very successful: I developed a Polish sketch grammar adapted for the GDPL group, showed them examples of extracted MWE results, and demonstrated how this approach could be directly integrated into their daily work. And although the more ambitious goal of automatic computation of the semantic compositionality of collocations was less successful, my appetite has been whetted and I hope to pursue this problem further soon, hopefully in further collaboration with the PWNG. Also, I hope that by visiting and discussing with three different leading lexicography-related research groups in Poland, I have also helped them develop potential further colloboration with each other. In any case, I have certainly learned a lot.

I include below screenshots of my Sketch Engine results for an example word ("samochód" = "car"). I attach also my sketch grammar and an example (also for "samochód") of the collocations extracted in XML with the Sketch Engine Script. I do not include the adjusted HTML editor page, as this is internal to GDPL.



"samochód" Word Sketch - Google Chrome

"samochód" Word Sk×

https://beta.sketchengine.co.uk/bonito/run.cgi/wsketch?corpname=user%2Fcjvt02%2Fgramdev_pltenten12_sample&reload=&lemma=samochód&lpos=&minfreq=auto&mir

Sketch Engine

Send feedback corpus: plTenTen12 [sample] Dr. Cyprian Laskowski

samochód (noun) plTenTen12 [sample] freq = 15,185 (274.18 per million)

	rzecz_nad_przym	8,412	1.80
□	ciężarowy	413	10.49
□	osobowy	790	10.34
□	dostawczy	228	9.69
□	terenowy	131	8.64
□	służbowy	101	8.35
□	używać	194	7.80
□	luksusowy	68	7.77
□	zastępczy	56	7.49
□	zabytkowy	56	7.36
□	zaparkować	34	6.97
□	elektryczny	58	6.88
□	uszkodzony	33	6.88
□	sportowy	73	6.71
□	rajdowy	29	6.71
□	wyścigowy	26	6.57
□	nowy	323	6.56
□	własny	132	6.54
□	drogi	38	6.41
□	użytkowy	29	6.35
□	twój	94	6.24
□	prywatny	43	6.20
□	firmowy	27	6.19
□	napędzać	20	6.13
□	pożarniczy	19	6.13
□	kempingowy	18	6.09
			>>

	rzecz_pod_rzecz_d	4,811	2.40
□	wypożyczalnia	178	9.85
□	wynajem	155	9.26
□	naprawa	66	7.93
□	zakup	182	7.75
□	kupno	48	7.66
□	zderzenie	35	7.66
□	kierowca	83	7.64
□	parkować	32	7.59
□	wypożyczyć	35	7.50
□	oklejać	27	7.45
□	kradzież	33	7.36
□	model	134	7.32
□	producent	109	7.29
□	skup	26	7.26
□	sprzedaż	134	7.20
□	posiadacz	35	7.15
□	dealer	24	7.14
□	flota	27	7.06
□	fabryka	35	7.03
□	zaparkować	19	6.88
□	karoseria	18	6.87
□	właściciel	70	6.82
□	ubezpieczenie	64	6.79
□	leasing	21	6.79
□	myć	19	6.66
			>>

	rzecz_pod_czas_podm	1,532	1.60
□	zderzyć	30	8.88
□	wjechać	12	7.53
□	uderzyć	8	6.41
□	jeździć	11	6.21
□	jechać	13	5.96
□	kosztować	7	5.83
□	zatrzymać	11	5.78
□	stać	36	5.37
□	ciesać	6	5.23
□	sprawiać	9	5.03
□	pojawić	13	4.63
□	sprawdzać	6	4.58
□	kupować	6	4.49
□	poruszać	6	4.45
□	musieć	29	4.34
□	stawać	8	4.29
□	potrafić	7	4.14
□	posiadać	24	4.11
□	uczestniczyć	7	4.02
□	czekać	6	3.97
□	być	425	3.93
□	zostać	59	3.72
□	wyglądać	6	3.50
□	znajdować	13	3.48
□	maić	14	3.38
			>>

	rzecz_pod_czas_b	1,233	2.40
□	parkować	21	8.54
□	zaparkować	19	8.52
□	naprawiać	13	7.93
□	wynająć	26	7.73
□	kupić	60	7.72
□	kupować	52	7.65
□	zostawić	19	7.62
□	ukraść	8	7.37
□	wypożyczyć	12	7.18
□	zatrzymywać	11	7.04
□	wjechać	7	6.97
□	zostawiać	9	6.95
□	wyprzedzać	6	6.69
□	sprzedać	13	6.46
□	zakupić	18	6.38
□	nabywać	8	6.30
□	myć	6	6.26
□	zatrzymać	12	5.97
□	produkować	13	5.80
□	pozostawić	7	5.62
□	jeździć	6	5.42
□	sprzedawać	8	5.29
□	posiadać	44	5.00
□	nabyć	8	4.98
□	prowadzić	63	4.80
			>>

	rzecz_nad_rzecz_d	912	0.50
□	marka	249	9.08
□	volvo	9	7.72
□	BMW	10	7.60
□	Suzuki	6	7.31
□	ford	6	7.13
□	segment	13	6.83
□	ratownictwo	8	6.63
□	Leszek	6	5.88
□	klasa	36	5.41
□	straż	10	5.40
□	typ	33	4.52
□	świat	9	3.56
□	uczestnik	6	2.41
□	rok	30	1.70
□	firma	7	0.75
□	osoba	7	0.55
			>>

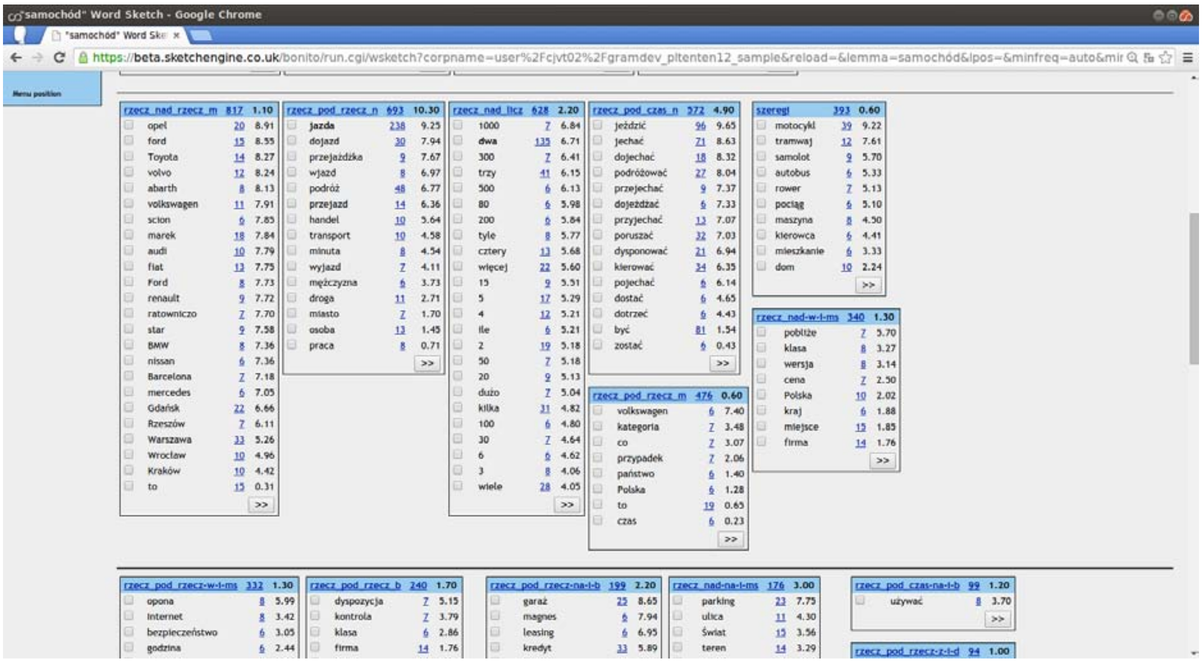
	rzecz_nad_rzecz_m	817	1.10
□	opel	20	8.91
□	ford	15	8.55

	rzecz_pod_rzecz_n	693	10.30
□	jazda	238	9.25
□	dojazd	30	7.94

	rzecz_nad_licz	628	2.20
□	1000	7	6.84
□	dwa	135	6.71

	rzecz_pod_czas_n	572	4.90
□	jeździć	96	9.65
□	jechać	71	8.63

	szeregi	393	0.60
□	motocykl	39	9.22
□	tramwaj	12	7.61



"samochód" Word Sketch - Google Chrome

"samochód" Word Sk×

https://beta.sketchengine.co.uk/bonito/run.cgi/wsketch?corpname=user%2Fcjvt02%2Fgramdev_pltenten12_sample&reload=&lemma=samochód&lpos=&minfreq=auto&mir

	rzecz_nad_rzecz_m	817	1.10
□	opel	20	8.91
□	ford	15	8.55
□	Toyota	14	8.27
□	volvo	12	8.24
□	abarth	8	8.13
□	volkswagen	11	7.91
□	scion	6	7.85
□	marek	18	7.84
□	audi	10	7.79
□	fiat	13	7.75
□	Ford	8	7.73
□	renault	9	7.72
□	ratowniczo	7	7.70
□	star	9	7.58
□	BMW	8	7.36
□	nissan	6	7.36
□	Barcelona	7	7.18
□	mercedes	6	7.05
□	Gdańsk	22	6.66
□	Rzeszów	7	6.11
□	Warszawa	33	5.26
□	Wrocław	10	4.96
□	Kraków	10	4.42
□	to	15	0.31
			>>

	rzecz_pod_rzecz_n	693	10.30
□	jazda	238	9.25
□	dojazd	30	7.94
□	przejażdżka	9	7.67
□	wjazd	8	6.97
□	podróż	48	6.77
□	przejazd	14	6.36
□	handel	10	5.64
□	transport	10	4.58
□	minuta	8	4.54
□	wyjazd	7	4.11
□	mężczyzna	6	3.73
□	droga	11	2.71
□	miasto	7	1.70
□	osoba	13	1.45
□	praca	8	0.71
			>>

	rzecz_nad_licz	628	2.20
□	1000	7	6.84
□	dwa	135	6.71
□	300	7	6.41
□	trzy	41	6.15
□	500	6	6.13
□	80	6	5.98
□	200	6	5.84
□	tyle	8	5.77
□	cztery	13	5.68
□	więcej	22	5.60
□	15	9	5.51
□	5	17	5.29
□	4	12	5.21
□	ile	6	5.21
□	2	19	5.18
□	50	7	5.18
□	20	9	5.13
□	dużo	7	5.04
□	kilka	31	4.82
□	100	6	4.80
□	30	7	4.64
□	6	6	4.62
□	3	8	4.06
□	wiele	28	4.05
			>>

	rzecz_pod_czas_n	572	4.90
□	jeździć	96	9.65
□	jechać	71	8.63
□	dojechać	18	8.32
□	podróżować	27	8.04
□	przejechać	9	7.37
□	dojeżdżać	6	7.33
□	przyjechać	13	7.07
□	poruszać	32	7.03
□	dysponować	21	6.94
□	kierować	34	6.35
□	pojechać	6	6.14
□	dostać	6	4.65
□	dotrzeć	6	4.43
□	być	81	1.54
□	zostać	6	0.43
			>>

	rzecz_pod_rzecz_m	476	0.60
□	volkswagen	6	7.40
□	kategoria	7	3.48
□	co	7	3.07
□	przypadek	7	2.06
□	państwo	6	1.40
□	Polska	6	1.28
□	to	19	0.65
□	czas	6	0.23
			>>

	szeregi	393	0.60
□	motocykl	39	9.22
□	tramwaj	12	7.61
□	samolot	9	5.70
□	autobus	6	5.33
□	rower	7	5.13
□	pociąg	6	5.10
□	maszyna	8	4.50
□	kierowca	6	4.41
□	mieszkanie	6	3.33
□	dom	10	2.24
			>>

	rzecz_nad-w-l-ms	340	1.30
□	pobliże	7	5.70
□	klasa	8	3.27
□	wersja	8	3.14
□	cena	7	2.50
□	Polska	10	2.02
□	kraj	6	1.88
□	miejsce	15	1.85
□	firma	14	1.76
			>>

	rzecz_pod_rzecz-w-l-ms	332	1.30
□	opona	8	5.99
□	internet	8	3.42
□	bezpieczeństwo	6	3.05
□	godzina	6	2.44

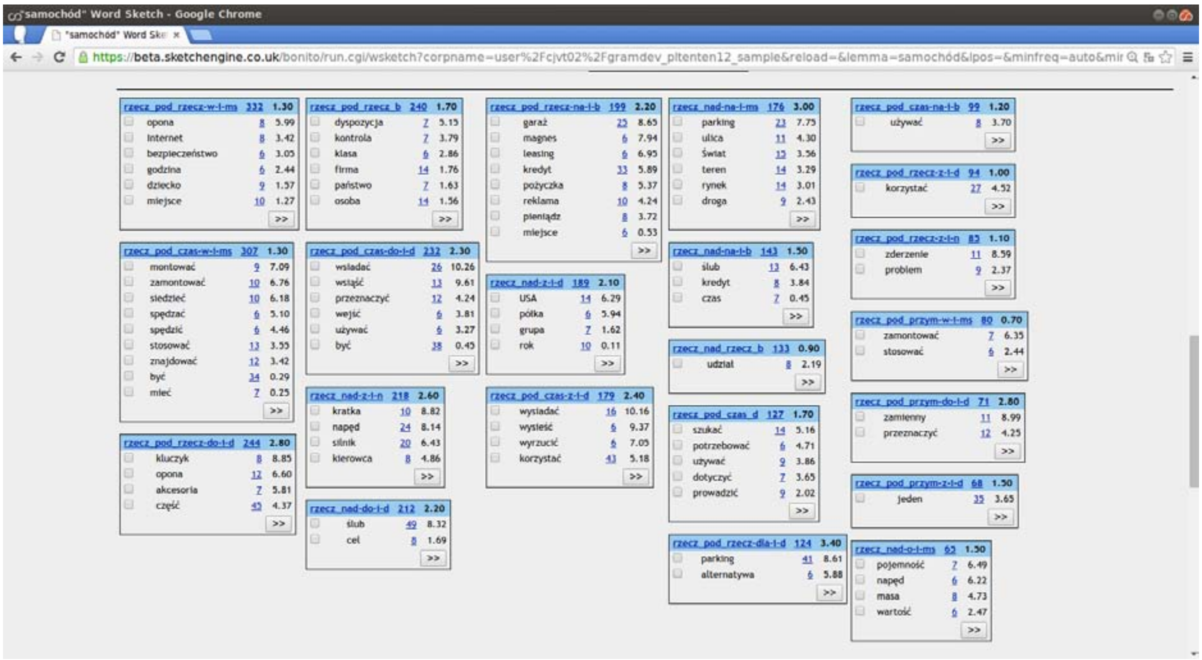
	rzecz_pod_rzecz_b	240	1.70
□	dyspozycja	7	5.15
□	kontrola	7	3.79
□	klasa	6	2.86
□	firma	14	1.76

	rzecz_pod_rzecz-na-l-b	199	2.20
□	garaż	25	8.65
□	magnes	6	7.94
□	leasing	6	6.95
□	kredyt	33	5.89

	rzecz_nad-na-l-ms	176	3.00
□	parking	23	7.75
□	ulica	11	4.30
□	Świat	15	3.56
□	teren	14	3.29

	rzecz_pod_czas-na-l-b	99	1.20
□	używać	8	3.70
			>>

	rzecz_pod_rzecz-z-l-d	94	1.00



"samochód" Word Sketch - Google Chrome

"samochód" Word Ske ×

https://beta.sketchengine.co.uk/bonito/run.cgi/wsketch?corpname=user%2Fcjvt02%2Fgramdev_pltenten12_sample&reload=&lemma=samochód&lpos=&minfreq=auto&mir

rzecz_pod_rzecz-w-l-ms	332	1.30
□ opona	8	5.99
□ internet	8	3.42
□ bezpieczeństwo	6	3.05
□ godzina	6	2.44
□ dziecko	9	1.57
□ miejsce	10	1.27
		>>

rzecz_pod_czas-w-l-ms	307	1.30
□ montować	9	7.09
□ zamontować	10	6.76
□ siedzieć	10	6.18
□ spędzać	6	5.10
□ spędzić	6	4.46
□ stosować	13	3.55
□ znajdować	12	3.42
□ być	34	0.29
□ mieć	7	0.25
		>>

rzecz_pod_rzecz-do-l-d	244	2.80
□ kluczyk	8	8.85
□ opona	12	6.60
□ akcesoria	7	5.81
□ część	45	4.37
		>>

rzecz_pod_rzecz_b	240	1.70
□ dyspozycja	7	5.15
□ kontrola	7	3.79
□ klasa	6	2.86
□ firma	14	1.76
□ państwo	7	1.63
□ osoba	14	1.56
		>>

rzecz_pod_czas-do-l-d	232	2.30
□ wsiadać	26	10.26
□ wsiąść	13	9.61
□ przeznaczyć	12	4.24
□ wejść	6	3.81
□ używać	6	3.27
□ być	38	0.45
		>>

rzecz_nad-z-l-n	218	2.60
□ kratka	10	8.82
□ napęd	24	8.14
□ silnik	20	6.43
□ kierowca	8	4.86
		>>

rzecz_nad-do-l-d	212	2.20
□ ślub	49	8.32
□ cel	8	1.69
		>>

rzecz_pod_rzecz-na-l-b	199	2.20
□ garaż	25	8.65
□ magnes	6	7.94
□ leasing	6	6.95
□ kredyt	33	5.89
□ pożyczka	8	5.37
□ reklama	10	4.24
□ pieniądz	8	3.72
□ miejsce	6	0.53
		>>

rzecz_nad-z-l-d	189	2.10
□ USA	14	6.29
□ półka	6	5.94
□ grupa	7	1.62
□ rok	10	0.11
		>>

rzecz_pod_czas-z-l-d	179	2.40
□ wysiadać	16	10.16
□ wysieść	6	9.37
□ wyrzucić	6	7.05
□ korzystać	43	5.18
		>>

rzecz_nad-na-l-ms	176	3.00
□ parking	23	7.75
□ ulica	11	4.30
□ Świat	15	3.56
□ teren	14	3.29
□ rynek	14	3.01
□ droga	9	2.43
		>>

rzecz_nad-na-l-b	143	1.50
□ ślub	13	6.43
□ kredyt	8	3.84
□ czas	7	0.45
		>>

rzecz_nad_rzecz_b	133	0.90
□ udział	8	2.19
		>>

rzecz_pod_czas_d	127	1.70
□ szukać	14	5.16
□ potrzebować	6	4.71
□ używać	9	3.86
□ dotyczyć	7	3.65
□ prowadzić	9	2.02
		>>

rzecz_pod_rzecz-dla-l-d	124	3.40
□ parking	41	8.61
□ alternatywa	6	5.88
		>>

rzecz_pod_czas-na-l-b	99	1.20
□ używać	8	3.70
		>>

rzecz_pod_rzecz-z-l-d	94	1.00
□ korzystać	27	4.52
		>>

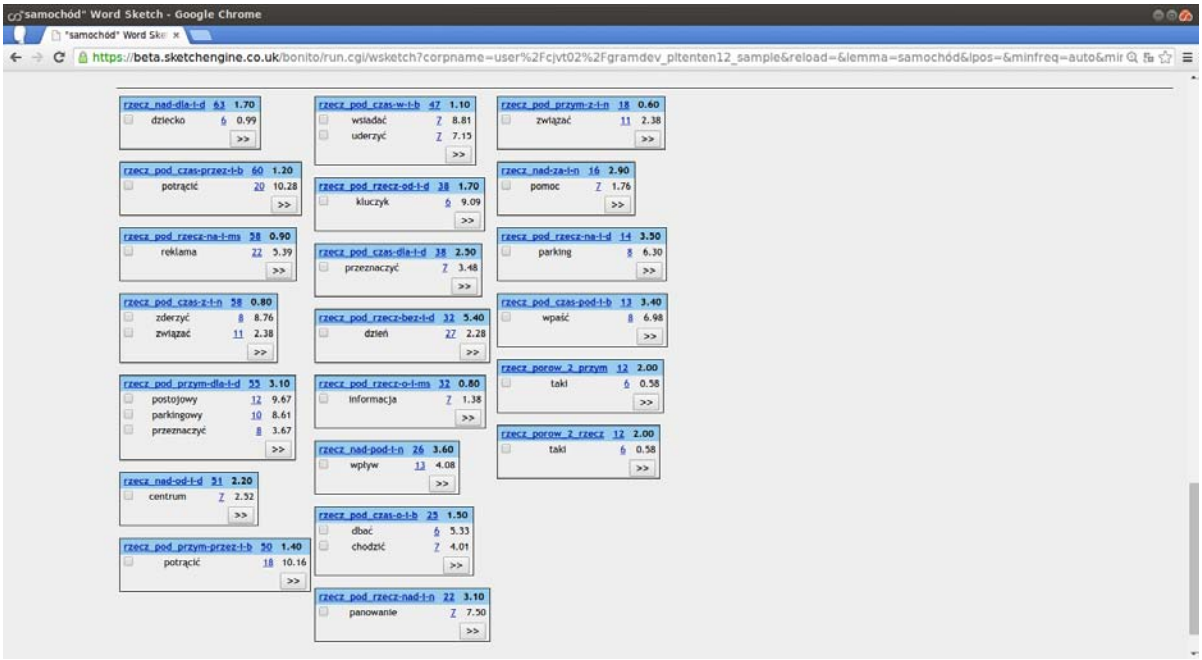
rzecz_pod_rzecz-z-l-n	85	1.10
□ zderzenie	11	8.59
□ problem	9	2.37
		>>

rzecz_pod_przym-w-l-ms	80	0.70
□ zamontować	7	6.35
□ stosować	6	2.44
		>>

rzecz_pod_przym-do-l-d	71	2.80
□ zamienny	11	8.99
□ przeznaczyć	12	4.25
		>>

rzecz_pod_przym-z-l-d	68	1.50
□ jeden	35	3.65
		>>

rzecz_nad-o-l-ms	65	1.50
□ pojemność	7	6.49
□ napęd	6	6.22
□ masa	8	4.73
□ wartość	6	2.47
		>>



"samochód" Word Sketch - Google Chrome

"samochód" Word Ske ×

https://beta.sketchengine.co.uk/bonito/run.cgi/wsketch?corpname=user%2Fcjvt02%2Fgramdev_pltenten12_sample&reload=&lemma=samochód&lpos=&minfreq=auto&mir

rzecz_nad-dla-l-d	63	1.70
□ dziecko	6	0.99
		>>

rzecz_pod_czas-przez-l-b	60	1.20
□ potrącić	20	10.28
		>>

rzecz_pod_rzecz-na-l-ms	58	0.90
□ reklama	22	5.39
		>>

rzecz_pod_czas-z-l-n	58	0.80
□ zderzyć	8	8.76
□ związać	11	2.38
		>>

rzecz_pod_przym-dla-l-d	55	3.10
□ postojowy	12	9.67
□ parkingowy	10	8.61
□ przeznaczyć	8	3.67
		>>

rzecz_nad-od-l-d	51	2.20
□ centrum	7	2.52
		>>

rzecz_pod_przym-przez-l-b	50	1.40
□ potrącić	18	10.16
		>>

rzecz_pod_czas-w-l-b	47	1.10
□ wsiadać	7	8.81
□ uderzyć	7	7.15
		>>

rzecz_pod_rzecz-od-l-d	38	1.70
□ kluczyk	6	9.09
		>>

rzecz_pod_czas-dla-l-d	38	2.50
□ przeznaczyć	7	3.48
		>>

rzecz_pod_rzecz-bez-l-d	32	5.40
□ dzień	27	2.28
		>>

rzecz_pod_rzecz-o-l-ms	32	0.80
□ informacja	7	1.38
		>>

rzecz_nad-pod-l-n	26	3.60
□ wpływ	13	4.08
		>>

rzecz_pod_czas-o-l-b	25	1.50
□ dbać	6	5.33
□ chodzić	7	4.01
		>>

rzecz_pod_rzecz-nad-l-n	22	3.10
□ panowanie	7	7.50
		>>

rzecz_pod_przym-z-l-n	18	0.60
□ związać	11	2.38
		>>

rzecz_nad-za-l-n	16	2.90
□ pomoc	7	1.76
		>>

rzecz_pod_rzecz-na-l-d	14	3.50
□ parking	8	6.30
		>>

rzecz_pod_czas-pod-l-b	13	3.40
□ wpaść	8	6.98
		>>

rzecz_porow_2_przym	12	2.00
□ taki	6	0.58
		>>

rzecz_porow_2_rzecz	12	2.00
□ taki	6	0.58
		>>

3. Links

- Sketch Engine: http://www.sketchengine.co.uk/
- Slovene sketch grammar: https://www.sketchengine.co.uk/documentation/raw-attachment/wiki/SKEW-3/Program/Krek_SKEW-3.pdf
- Narodowy korpus języka polskiego (National Corpus of Polish): http://nkjp.pl/index.php
- Wielki słownik języka polskiego (Great Dictionary of the Polish Language): http://www.wsjp.pl/
- Instytut języka polskiego (Polish Language Institute): https://www.ijp-pan.krakow.pl/
- Zespół Inżynierii Lingwistycznej (Linguistic Engineering Group): http://zil.ipipan.waw.pl/
- Słowosieć (plWordnet): http://plwordnet.pwr.wroc.pl/wordnet