

Jakich danych potrzebujemy?

Piotr Pęzik, Uniwersytet Łódzki, pezik@uni.lodz.pl

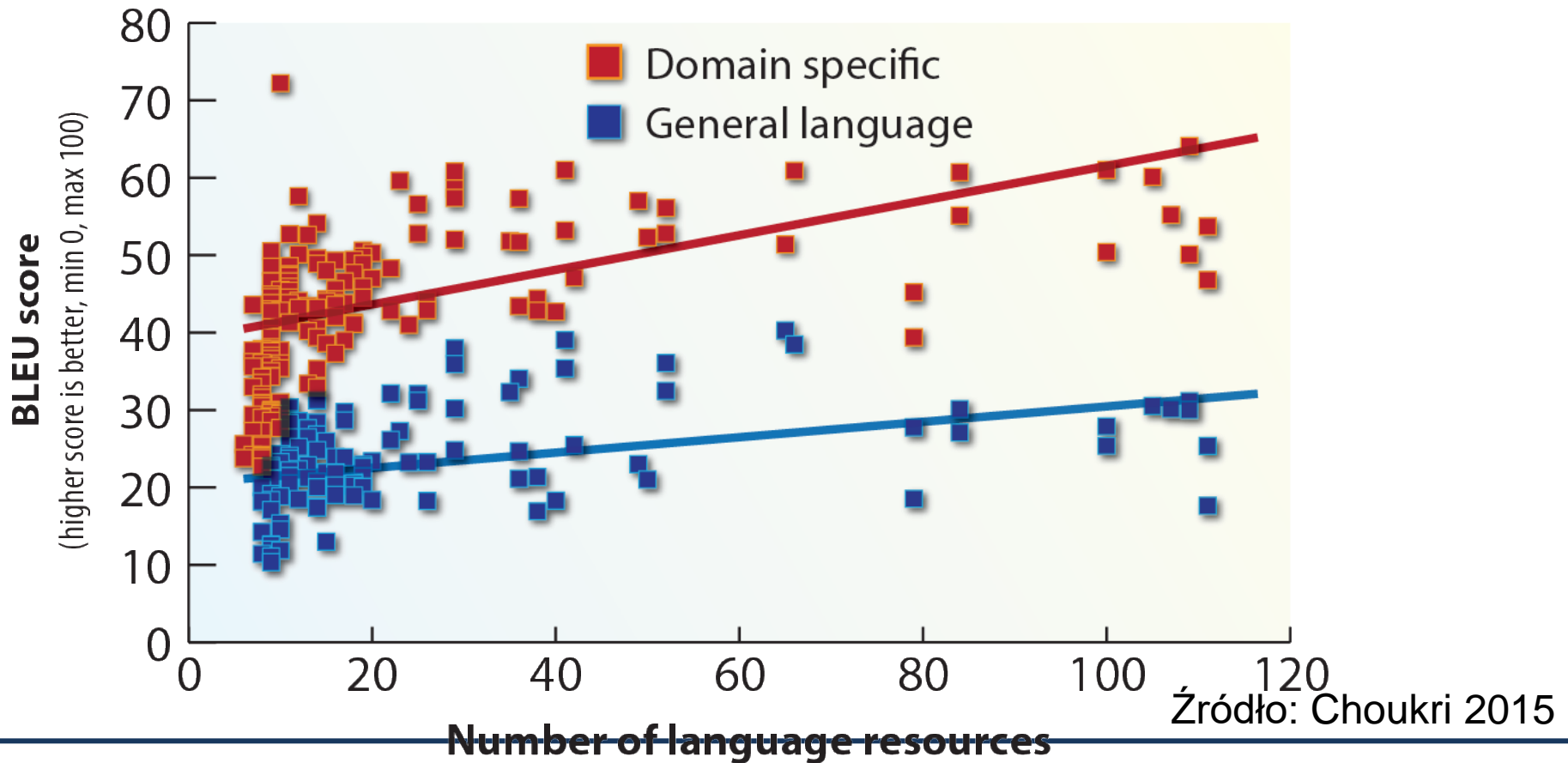
Dlaczego potrzebujemy danych?



- Statystyczne systemy TM są ‘trenowane’ na danych
- Im więcej dobrych, reprezentatywnych danych, tym większa szansa na poprawę jakości TM



Impact of number of language resources on Statistical MT quality



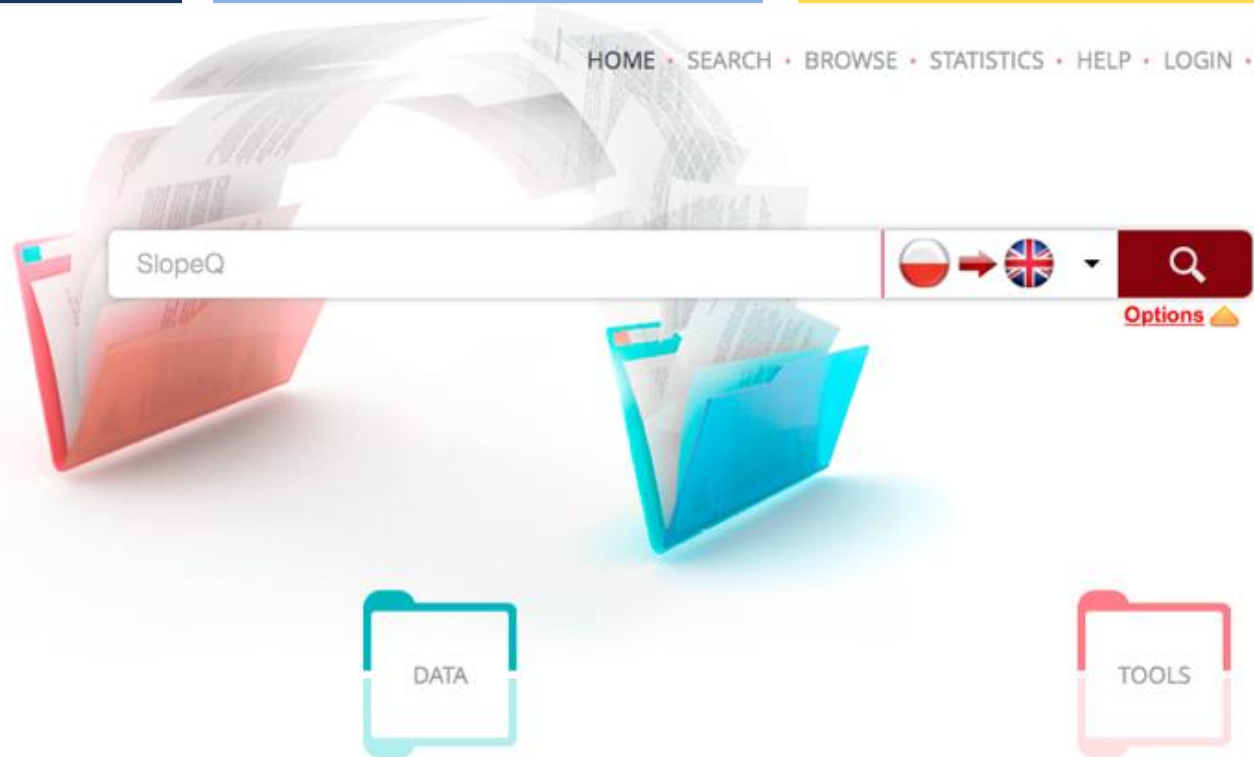
Źródło: Choukri 2015

Number of language resources

- Pamięci tłumaczeniowe
- Korpusy i kolekcje równoległe (strony WWW, transkrypcje tłumaczonych wystąpień, dokumentacja wewnętrzna)
- Teksty porównywalne (częściowe tłumaczenia i adaptacje)
- Słowniki elektroniczne

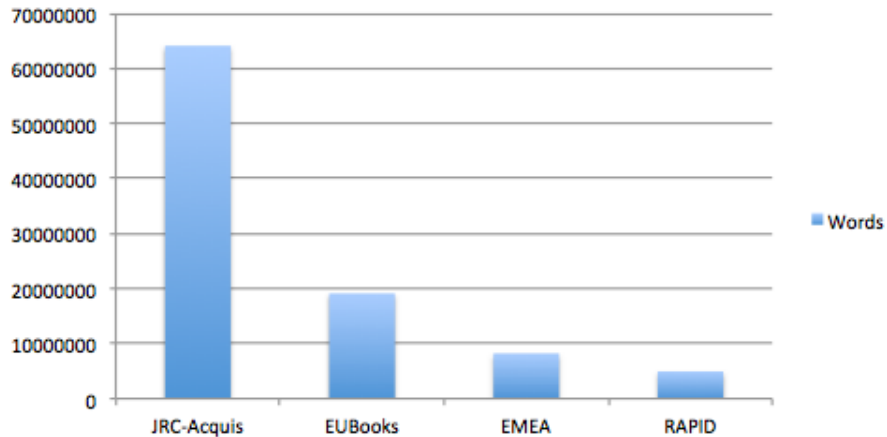


- Dużych i zróżnicowanych (reprezentatywnych i zbilansowanych)
- Dobrej jakości
- Aktualnych

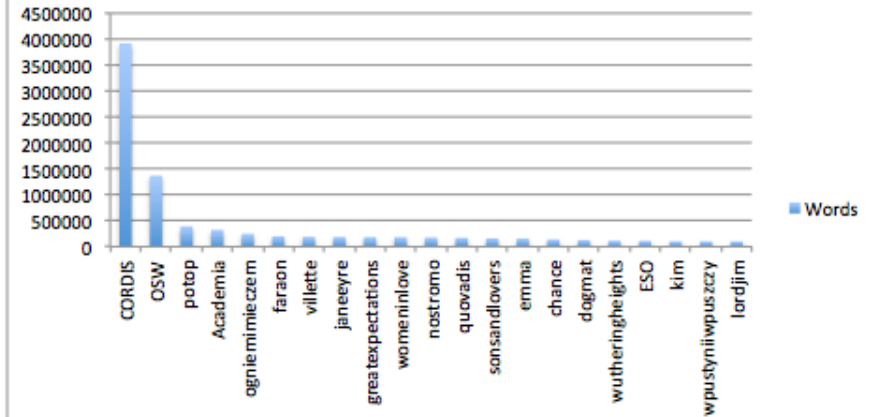


- <http://paralela.clarin-pl.eu>
- Ok. 262 mln słów w części polskiej, 10,8 mln segmentów

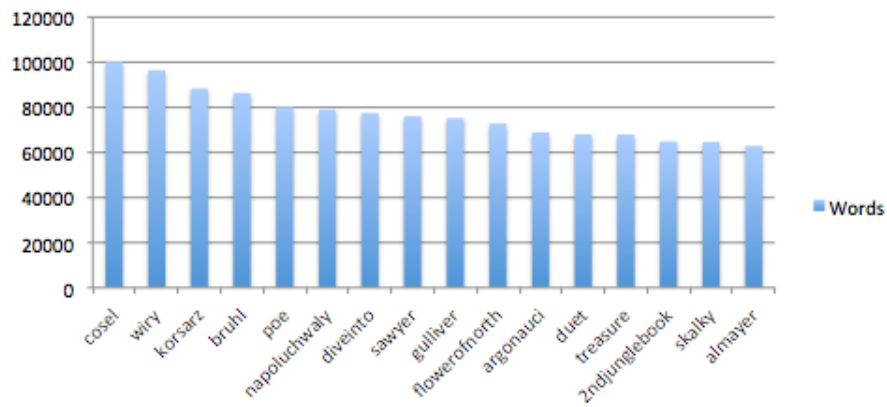
Sources 1..5



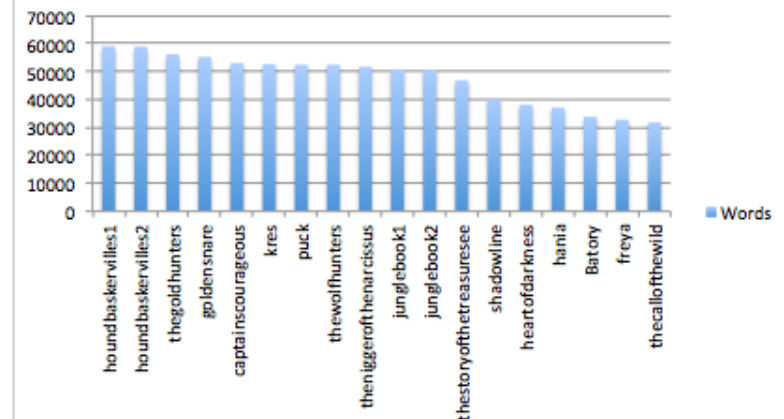
Sources 6..26

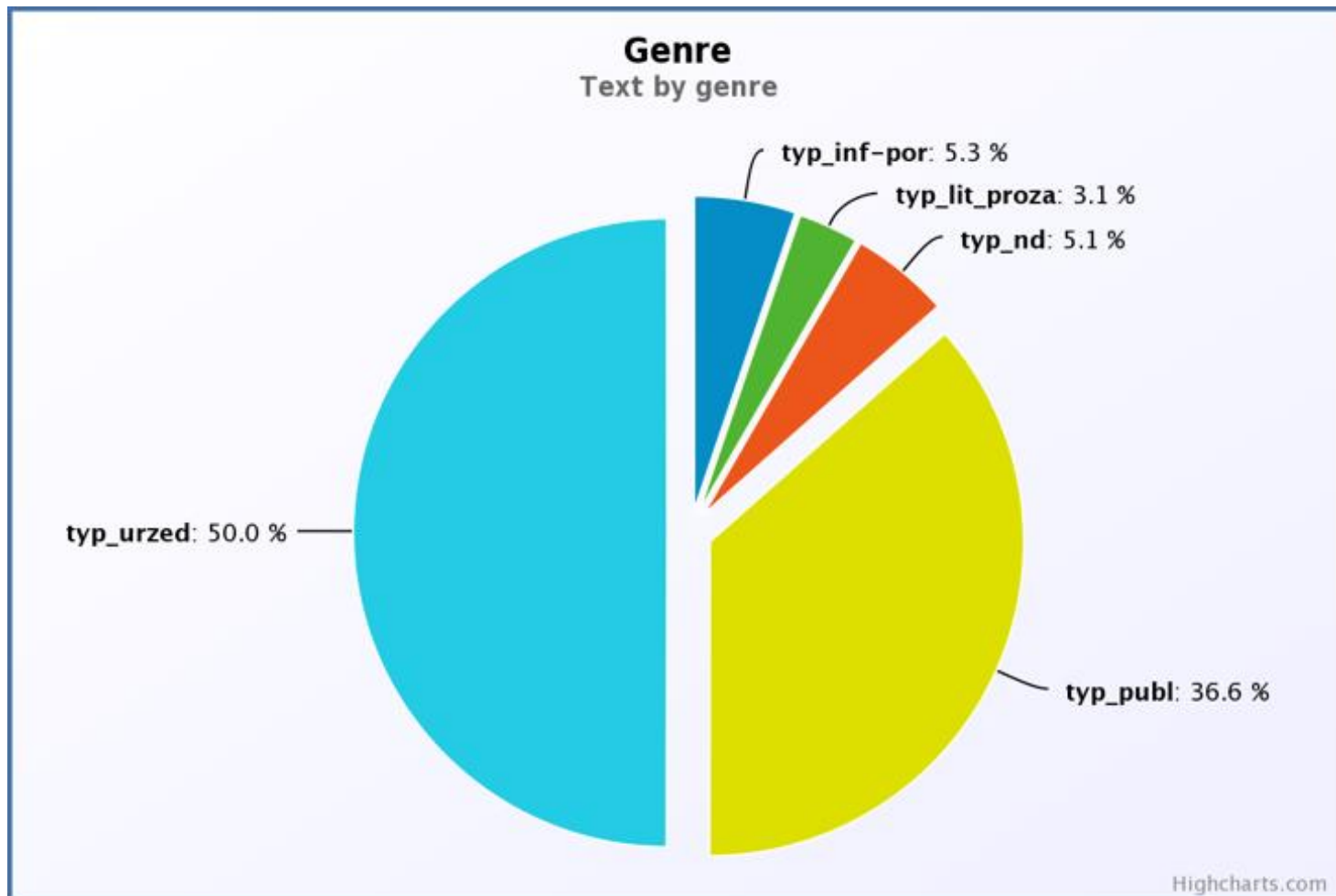


Sources 27..42



Sources 43..60







Alignment type
Segments by alignment type



Highcharts.com

5 To rzekłszy , chwycił za kark pana Andrzeja , a pan Andrzej chwycił jego ;
lecz poprzednio już dwaj młodzi Kiemlicze , Kosma i Damian , podnieśli
się z ławy , sięgając rozczochranemi głowami aż do pułapu i Kosma spytał
: — Ociec , **prać** ?

Potop 

First , however , the two young Kiemliches , Kosma and Damian , had
risen from the bench , almost touching the ceiling with their bushy heads ,
and Kosma asked,-- " Shall we pound , father ? "

The Deluge 

<http://paralela.clarin-pl.eu>



- W TM poza ‘dokładnością’ (ang. accuracy) ważna jest ‘płynność’ (‘fluency’)
- Płynność jako ‘idiomatyczność tłumaczenia’
- Idiomatic translation one “which has the same meaning as the source language, but is expressed in the natural form of the receptor language” and in which “the meaning not the form is retained” (Larson, 1984: 11), (cf. Beekman, Callow, 1974).
- To oznacza m.in. odpowiednie tłumaczenie jednostek frazeologicznych

Kill two birds with one stone



W korpusie <http://monitorcorpus.com> idiom ten występuje zaledwie 116 razy na 1827 mln. słów.















Facets Export

3 results. *Your criteria.*

een buying a boat or a house , you could	kill two birds with one stone
The pursuit could	kill two birds with one stone
ned from a cold training ride , she would	kill two birds with one stone
A digital AE-1 could	kill two birds with one stone
g when you need her ? Hester decides to	kill two birds with one stone
's Composure Collection that would also	kill two birds with one stone
	Kill two birds with one stone
stable smart dumbbells gave a chance to	kill two birds with one stone
You may want to	kill two birds with one stone
Combine a visit with a road trip and)	kill two birds with one stone

Dwie pieczenie przy jednym ogniu



4	How do we kill two birds with one stone ? <i>Bottoms Up</i> 	Jak upiec dwie pieczenie na jednym ogniu ? <i>Bottoms Up</i> 
5	We can kill two birds with one stone . <i>The Bag or the Bat</i> 	Upieczemy dwie pieczenie na jednym ogniu . <i>The Bag or the Bat</i> 
7	I figured I could kill two birds with one stone . <i>Dance with Somebody</i> 	Zdał em sobie sprawę , że mogę upiec dwie pieczenie na jednym ogniu . <i>Dance with Somebody</i> 
8	You can kill two birds with one stone , right ? <i>Ninja Scroll</i> 	Przyłącz się do mnie , a upieczesz dwie pieczenie na jednym ogniu . <i>Ninja Scroll</i> 
9	Owner knew he had bad tenants , wanted to kill two birds with one stone ? <i>Pretty Persuasion</i> 	Właściciel wiedział , że ma złych lokatorów i postanowił upiec dwie pieczenie na jednym ogniu . <i>Pretty Persuasion</i> 
10	Since she 'll be your partner in this venture , we do require one meeting with her anyway , so kill two birds with one stone . <i>Meet Bill</i> 	Skoro będzie pana współnikiem i tak wymagamy jednego spotkania . <i>Meet Bill</i> 
11	Since she 'll be your partner in this venture , we do require one meeting with her anyway , so kill two birds with one stone . <i>Meet Bill</i> 	Upieczemy dwie pieczenie przy jednym ogniu . <i>Meet Bill</i> 

- Ok. 50 segmentów w korpusie Paralela
- Ok. 30 tłumaczeń na “upiec dwie pieczenie przy jednym ogniu”
- Można liczyć na wychwycenie ekwiwalencji przez system TM



<p>as the benefit of the doubt .</p> <p><i>Resistance Is Futile</i> </p>	<p>Jesteś jedną z nas , Doakes jest jednym z nas , zasłużył na zaufanie .</p> <p><i>Resist</i></p>
<p>checks have been made , growers shall be given the benefit of the doubt .</p> <p><i>Porządek prawny UE</i> </p>	<p>Po przeprowadzeniu wszystkich kontroli wątpliwości rozstrzygnięć na korzyść plantatorów .</p> <p><i>Acquis co</i></p>
<p>ing to give him the benefit of the doubt ?</p> <p><i>Suspicious Minds</i> </p>	<p>Więc ciągle chcesz mu dać " domniemanie niewinności " .</p> <p><i>Su</i></p>
<p>ing to give him the benefit of the doubt .</p> <p><i>Suspicious Minds</i> </p>	<p>Zamierzam podarować mu</p> <p><i>Su</i></p>
<p>ing to give him the benefit of the doubt .</p> <p><i>Suspicious Minds</i> </p>	<p>" domniemanie niewinności " .</p> <p><i>Su</i></p>
<p>want to give her the benefit of the doubt .</p> <p><i>Three for the Road</i> </p>	<p>Wybacz , że chcę jej dać kredyt zaufania !</p> <p><i>Thro</i></p>

- 109 segmentów, bardzo niestabilna ekwiwalencja
- Czego z takich danych może się nauczyć system TM?

19 Just trying to **kill two birds with one stone** .

Way Out West 

Tak , jakby spróbować zabić dwa ptaki przy pomocy jednego kamienia .

Way Out West 

3 Such a measure will enable us to **kill several birds with one stone** : it is an effective instrument against speculation and the revenue will enable us to address government deficits and the funding for pressing social projects (the environment , development aid , infrastructure projects , etc .) .

Parlament Europejski 

Instrument ten umożliwi nam złapanie kilku srok za ogon : jest to skuteczne narzędzie przeciwko spekulacji , a dochody jakie przyniesie umożliwią nam zlikwidowanie deficytów publicznych oraz sfinansowanie pilnych projektów społecznych (w dziedzinie środowiska , pomocy rozwojowej , infrastruktury itp .) .

European Parliament 

10 Thought I 'd **kill two birds with one stone** , you know .

Notting Hill 

Dwa grzyby w barszcz .

Notting Hill 

yszukiwarka korpusowa PELCRA dla danych NKJP (Cytowanie)

okować** produkt**

maks. odstęp: 1 Zachowaj szyk: Wyniki: 100

zaawansowane Podkorpus: Cały

<< Poprzednie

Następne >>

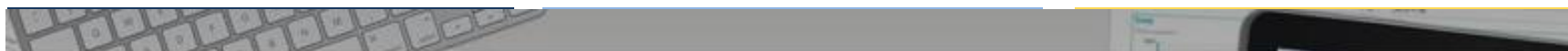
Pomoc

zeszukiwany zbiór zawiera 1,524,696,745 słów. Znaleziono 15 akapitów pasujących do zapytania w 0.027s. Bieżąca strona zawiera 15 przykładów z 14 różnych tekstów.

stać wyłącznie rynkiem lokalnym, na którym będą	lokowali swoje produkty	przedsiębiorcy z Unii Europejskiej, ale stąd już w	Sprawozdanie stenog...
Wprawdzie nadal zdecydowana większość składek jest	lokowana w produktach	bezpiecznych, jednak, coraz więcej inwestorów	Gazeta Ubezpieczeni...
Unii Europejskiej do Stanów Zjednoczonych w zakresie	lokowania produktów	(ang. product placement). Za oceanem ta usługa jest	Dziennik Internautó...
działalność w strefie i poza nią, np. poprzez	lokowanie końcowego produktu	w strefie i w konsekwencji całego zysku, nie	Sprawozdanie stenog...
jej na gorsze pozycje. Coraz trudniejsza sytuacja z	lokowaniem produktów	na rynku powoduje powrót producentów do korzystania z	Zagraniczne inwestyc...
cechą tego projektu jest to, że ułatwi on	lokowanie pewnych produktów	na nadwyżkowym rynku żywnościowym, a także	
Producenci w USA zarabiają krocie biorąc kasę za	lokowanie produktów	w filmie. Zyski z dystrybucji to już tylko dodatek.	Usenet -- pl.rec.fi...
Nie będzie możliwe	lokowanie produktów	podczas programów informacyjnych, poradnikowych i dla	Dziennik Internautó...
leżeć mogła, bo niektóre sklepy - kierując się nazwą -	lokują ten produkt	na stoisku rybnym). Paszтет łososiowy jest bowiem	Dziennik Polski
napływ nowych środków, jakie klienci regularnie	lokują w produktach	inwestycyjno-ochronnych. Według danych za pierwszy	Gazeta Ubezpieczeni...
ośmioma subfunduszami, które całość aktywów	lokują w produktach	inwestycyjnych znajdujących się w międzynarodowej	Gazeta Ubezpieczeni...
2. inwestowanie w osiem subfunduszy, które całość aktywów	lokują w produktach	inwestycyjnych znajdujących się w międzynarodowej	Gazeta Ubezpieczeni...
3. tylna opona. Ma aż 240 mm szerokości. Triumph również	lokuje nowy produkt	w kategorii cruiserów, ale to chyba trochę zbyt	Dziennik Bałtycki
4. że dzisiaj ta firma dużo eksportuje za granicę, dużo	lokuje produktów	na rynku niemieckim, na rynku duńskim, na rynkach	
5. Krótko: Przeciętna jakość obrazu, ale za to niska cena	lokuje te produkty	w kręgu zainteresowań osób, które bardzo chcą pokazać	Enter

Tylko kilka wystąpień “lokowania produktu” w NKJP do 2010 r.

<http://nkjp.uni.lodz.pl>



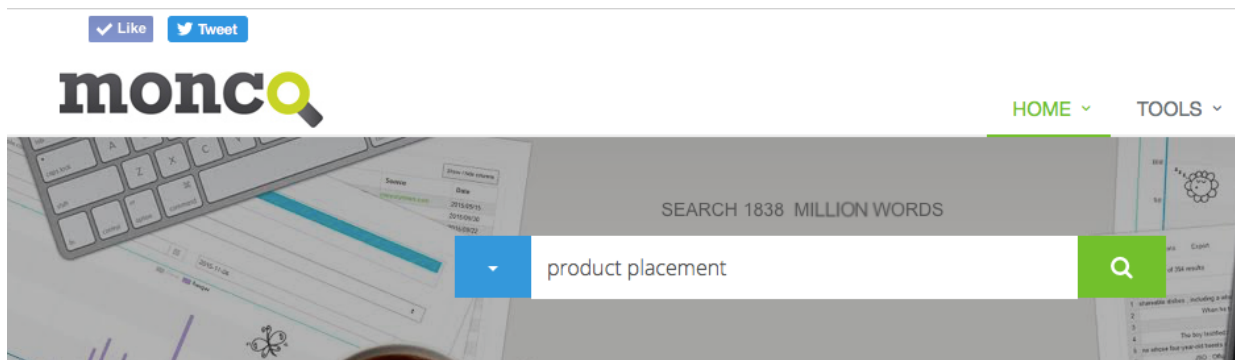
Konkordancje Podsumowanie Fasety Pliki

Wyniki od 1 do 20 z 804. Kryteria wyszukiwania. Show / hide columns

#			Źródło	Data
1	. Ustawa utrzymuje co do zasady zakaz tzw	lokowania produktów	wprost.pl	4/4/2011
2	-Ustawa utrzymuje co do zasady zakaz tzw	lokowania produktów	wprost.pl	4/4/2011
3	w końcu odnalazł em , ale dzięki zupełnie innej wyszukiwarce	lokowania produktu	gazeta.pl	13/2/2012
4	Tego typu	lokowanie produktów	polygamia.pl	1/12/2011
5	Uwaga , artykuł zawiera	lokowanie produktów	iswinoujscie.pl	26/5/2012
6	zupełny przypadek , ale tak właśnie wygląda reklama poprzez	lokowanie produktu	gosc.pl	19/10/2015
7	Autor : bober88 23 / 10 / 2011 23 : 19 Uwaga : artykuł zawiera	lokowanie produktu	andrychow.org	23/10/2011
8	Audycja zawiera	lokowanie produktu	gazeta.pl	13/2/2012
9) Product placement	lokowanie produktu	halolodz.pl	5/3/2012
10	nni oznaczyć materiały reporterów paskiem * Audycja zawiera	lokowanie produktu	gazeta.pl	14/11/2012
11	foze ja sie nie znam ale w tv chociaz pisza ze audycja zawiera	lokowanie produktu	wp.pl	16/1/2013
12	Oczywiście takie gry będą zawierały	lokowanie produktu	antyweb.pl	21/1/2016
13	any czy wykorzystywany w telewizji branded content (tudzież	lokowanie produktu	wirtualnemedi.pl	13/10/2013
14	stał mało widoczny dopisek niewielką czcionką ; tekst zawiera	lokowanie produktu	gazeta.pl	4/10/2013
15	Tekst zawiera	lokowanie produktu	elubin.pl	16/8/2013
16		Lokowanie produktu	wirtualnemedi.pl	1/2/2012
17		Lokowanie produktu	forsal.pl	30/1/2014
18	Sporo czasu spędzili śmy omawiając zagadnienia związane z	lokowaniem produktów	wp.pl	2/2/2012
19	więcej i coraz bardziej zaawansowanych działań związanych z	lokowaniem produktu	wirtualnemedi.pl	1/2/2012

804 wystąpienia w MonCo (2010 — 2015)

<http://monco.frazeo.pl>

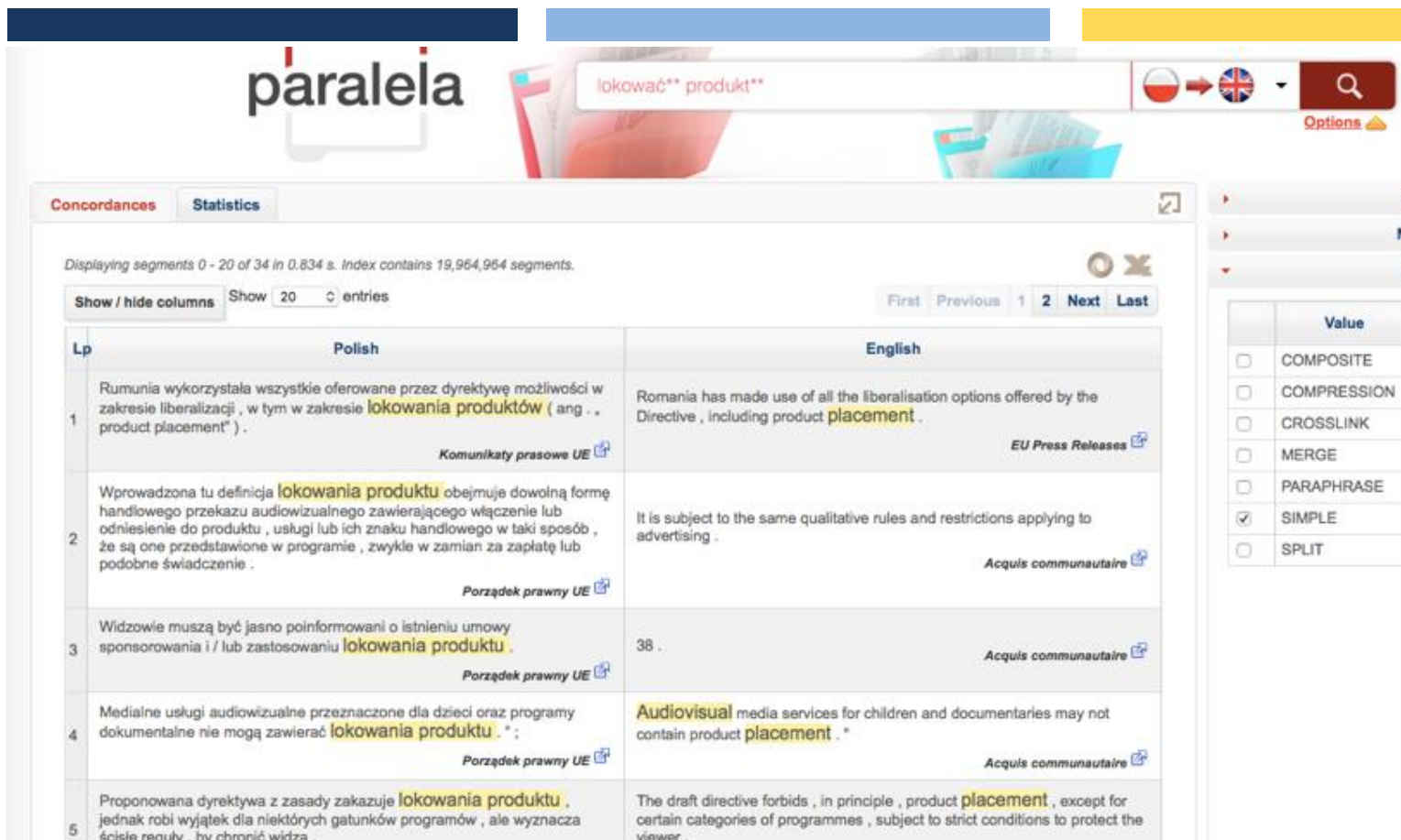


[Results](#)
[Summary](#)
[Facets](#)
[Export](#)

Showing 1 to 20 of 425 results. [Your criteria.](#)

#	Text snippet	Product placement
1	some tanning . ' ' Hey that 's racist ! ' he called back , ' Is this a	product placement ?
2	With a little help from some obvious LG	product placement , the couple got busy making cupcakes in the kitchen of thei...
3	imes ' A.O. Scott complained , that the movie's most egregious	product placement often makes it seem like a "90-minute commercial for Popey...
4	w album On Another Note , creating perhaps the finest piece of	product placement in our television history - more than able to serenade them .
5	CM was included in the windows and got the benefit of in-show	product placement as well .
6	hands her own social media following , charging up to \$ 500 per	product placement in Instagram posts .
7	0,000 followers on Instagram and can command up to \$ 500 for	product placement in posts Adding that she monitors her daughter 's exposure ...
8	CM was included in the windows and got the benefit of in-show	product placement as-well .
9	CM was included in the windows and got the benefit of in-show	product placement as-well .
10	tune in , which might go down as the most hilarious attempt at	product placement in broadcasting history .

425 wystąpień w korpusie monitorcorpus.com, product placement w mediach społecznościowych



The screenshot shows the Paralela interface with a search query "lokować** produkt**". The results are displayed in a table with columns for "Polish" and "English". The table contains 5 rows of concordance pairs. The interface also includes a search bar, navigation buttons (First, Previous, 1, 2, Next, Last), and a sidebar with a "Value" column and checkboxes for various options like COMPOSITE, COMPRESSION, CROSSLINK, MERGE, PARAPHRASE, SIMPLE, and SPLIT.

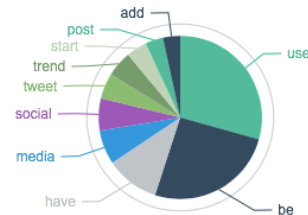
Lp	Polish	English
1	Rumunia wykorzystala wszystkie oferowane przez dyrektywę możliwości w zakresie liberalizacji , w tym w zakresie lokowania produktów (ang . „ product placement”) . <i>Komunikaty prasowe UE</i>	Romania has made use of all the liberalisation options offered by the Directive , including product placement . <i>EU Press Releases</i>
2	Wprowadzona tu definicja lokowania produktu obejmuje dowolną formę handlowego przekazu audiowizualnego zawierającego włączenie lub odniesienie do produktu , usługi lub ich znaku handlowego w taki sposób , że są one przedstawione w programie , zwykle w zamian za zapłatę lub podobne świadczenie . <i>Porządek prawny UE</i>	It is subject to the same qualitative rules and restrictions applying to advertising . <i>Acquis communautaire</i>
3	Widzowie muszą być jasno poinformowani o istnieniu umowy sponsorowania i / lub zastosowaniu lokowania produktu . <i>Porządek prawny UE</i>	38 . <i>Acquis communautaire</i>
4	Medialne usługi audiowizualne przeznaczone dla dzieci oraz programy dokumentalne nie mogą zawierać lokowania produktu . * ; <i>Porządek prawny UE</i>	Audiovisual media services for children and documentaries may not contain product placement . * <i>Acquis communautaire</i>
5	Proponowana dyrektywa z zasady zakazuje lokowania produktu , jednak robi wyjątek dla niektórych gatunków programów , ale wyznacza ścisłe reguły , by chronić widza .	The draft directive forbids , in principle , product placement , except for certain categories of programmes , subject to strict conditions to protect the viewer .

34 pary segmentów w Paraleli, głównie dzięki dobrej reprezentacji “Porządku prawnego UE”

Collocations

Analyzed: 2594 non-duplicated spans. Total spans for this query: 15565 .

Top 10 collocates



Search:

Showing 1 to 100 od 476 results. [Your criteria.](#)

#	Lemma	Frequency	Positions	Score	Paths
1	use	585	{-4=33, 5=9, -1=17, -2=451, 3=10, 4=8, ...	0.0054	{using the hashtag=308},{use the hashtag=78},{used th
2	be	512	{-5=55, 4=47, 3=61, 1=109, 2=46, -3=6...	0.0002	{hashtag is=59},{hashtag was=41},{are using the hasht
3	have	212	{3=31, 1=28, -3=34, -5=28, -4=29, 2=3...	0.0002	{hashtag has=21},{hashtag had=6},{hashtag that has=
4	media	135	{-3=19, -1=27, -4=25, -5=18, 5=21, 3=...	0.0035	{media hashtag=27},{hashtag on social media=13},{me
5	social	127	{-4=19, -2=27, -5=24, 4=20, 2=15, 3=1...	0.0052	{social media hashtag=27},{hashtag on social=13},{soc
6	tweet	102	{-2=5, -5=24, -3=32, 3=7, 4=4, 5=10, -...	0.0052	{tweet using the hashtag=6},{tweeted the hashtag=4},{
7	trend	101	{4=30, -3=2, 5=11, 2=21, 3=19, -2=3, -...	0.0147	{hashtag started trending=3},{hashtag to trend=3},{tr

Frazeologia „hashtagów”, wg monitorcorpus.com



- Pamięci tłumaczeniowe
- Przetłumaczone/ zlokalizowane strony
- Transkrypcje i nagrania tłumaczonych wystąpień
- Słowniki i domenowe bazy terminologiczne

```
<?xml version="1.0" encoding="UTF-8"?>
<xliff xmlns:None="urn:oasis:names:tc:xliff:document:1.2" version="1.2">
  <file source-language="pl" target-language="en" datatype="plaintext"
  original="bogactwo_42-45_kundzewicz_PL-richness_42-45_kundzewicz_ENalignment"
  <body>
    <trans-unit id="1">
      <source>Niezbędna jest adaptacja</source>
      <target>Mankind Must Adapt</target>
    </trans-unit>
    <trans-unit id="2">
      <source>Rozmowa Akademii ACADEMIA</source>
      <target>Briefly Speaking ACADEMIA</target>
    </trans-unit>
    <trans-unit id="3">
      <source>Wywiad z prof. Zbigniewem Kundzewiczem</source>
      <target>Interview with Prof. Zbigniew Kundzewicz</target>
    </trans-unit>
    <trans-unit id="4">
      <source>Niezbędna jest adaptacja</source>
      <target>Mankind Must Adapt</target>
    </trans-unit>
  </body>
</xliff>
```

Pamięci tłumaczeniowe (XLIFF, TMX)

Nauka online
portal magazynu
Polskiej Akademii Nauk **ACADEMIA**

EN PL FACEBOOK PA



Start

Na bieżąco

Polecamy

Teksty

Felietony

Autorzy

Academia



Możliwe jest zrównoleglenie automatyczne takich kolekcji

Telescope

From Wikipedia, the free encyclopedia


For other uses, see [Telescope \(disambiguation\)](#).

A **telescope** is an instrument that aids in the observation of remote objects by collecting [electromagnetic radiation](#) (such as [visible light](#)). The first known practical telescopes were invented in the [Netherlands](#) at the beginning of the 17th century, using glass lenses. They found use in terrestrial applications and astronomy.

Within a few decades, the [reflecting telescope](#) was invented, which used mirrors. In the 20th century many new types of telescopes were invented, including [radio telescopes](#) in the 1930s and [infrared telescopes](#) in the 1960s. The word *telescope* now refers to a wide range of instruments detecting different regions of the [electromagnetic spectrum](#), and in some cases other types of detectors.

The word "telescope" (from the [Ancient Greek](#) [τήλε](#), *tele* "far" and [σκοπεῖν](#), *skopein* "to look or see"; [τηλεσκόπος](#), *teleskopos* "far-seeing") was coined in 1611 by the Greek mathematician [Giovanni Demisiani](#) for one of [Galileo Galilei](#)'s instruments presented at a banquet at the [Accademia dei Lincei](#).^{[1][2][3]} In the *Starry Messenger*, Galileo had used the term "perspicillum".

Teleskop [edytuj]

 Ten artykuł dotyczy urządzenia optycznego. Zobacz też: [inne znaczenia tego słowa](#).

Teleskop (gr. *tēle-skópos* – daleko widzący) – narzędzie, które służy do obserwacji odległych obiektów poprzez zbieranie promieniowania [elektromagnetycznego](#) (np. światła widzialnego).

Pierwsze znane praktyczne teleskopy zostały skonstruowane przy użyciu szklanych soczewek w Holandii na początku XVII wieku przez [Hansa Lippersheya](#), a wkrótce potem przez [Galileusza](#) we Włoszech. Znalazły one zastosowanie w działaniach militarnych i w astronomii.

Przez wiele lat używano [zwierciadlanych teleskopów optycznych](#). W XX wieku opracowano nowe typy jak na przykład: [radioteleskop](#) w 1930 roku czy [teleskop podczerwony](#) w 1960 roku. Słowo teleskop odnosi się obecnie do szerokiej gamy instrumentów wykrywających różne zakresy widma elektromagnetycznego, a w niektórych przypadkach także do innych typów detektorów.

Możliwe przynajmniej częściowe zrównoleglenie takich tekstów



Wiele takich danych znajduje się w zasobach “sieci ukrytej”



www.deepwebtech.com



pezik@uni.lodz.pl