

СОСТАВЛЕНИЕ ПЕРЕЧНЯ
СОКРАЩЕННЫХ НАЗВАНИЙ ЯЗЫКОВ
В РАМКАХ ПРОЕКТА ДИГИТАЛИЗАЦИИ
«СЛОВАРЯ ПОЛЬСКОГО ЯЗЫКА»
С.Б.Линде

Иоанна Билиньска (Joanna Bilińska)

j.bilinska@uw.edu.pl

Варшавский университет, Кафедра формальной лингвистики,
Варшава, Польша

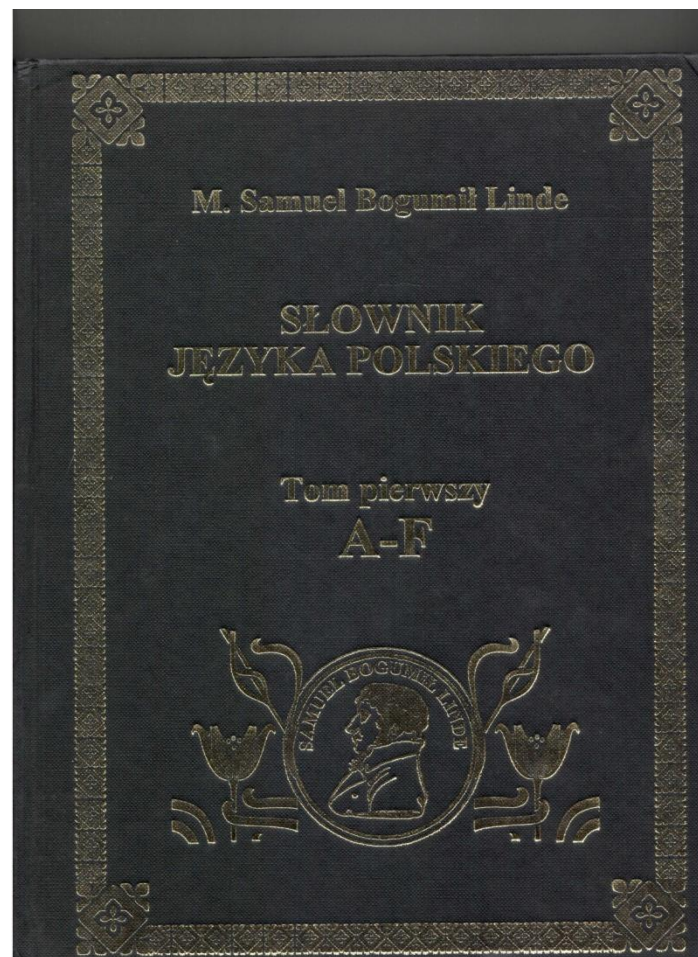
«Словарь польского языка»

- 1807-1814,
- Автор: Самуил Богумил Линде,
- Шеститомный,
- цифровая версия словаря, одна из наиболее популярных и часто используемых публикаций на <http://kpbc.umk.pl/dlibra>



<http://poliqarp.wbl.klf.uw.edu.pl/slownik-lindego>

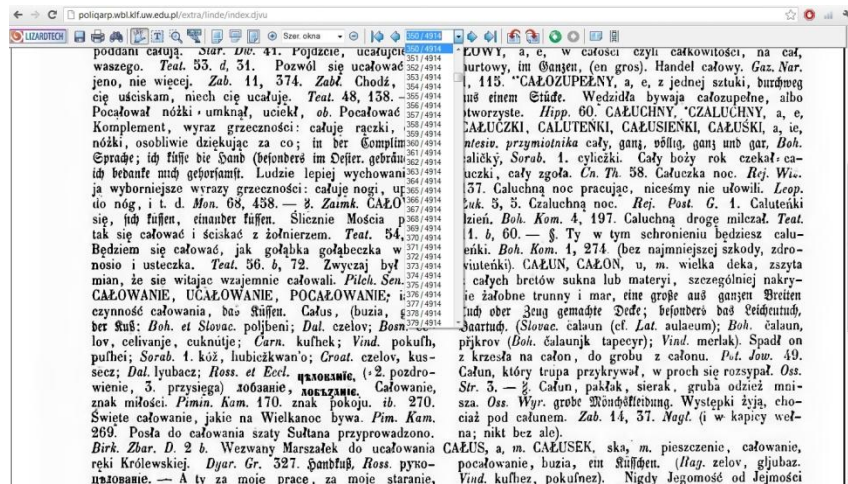
- в Кафедре формальной лингвистики,
- 2-е издание словаря Линде,
- с использованием технологии оптического распознавания текста (OCR),
- одновременно совмещенная с поисковой системой,



<http://poliqarp.wbl.klf.uw.edu.pl/slownik-lindego/> (2)

- Словарь как корпус текстов,
- около 7 миллионов сегментов
- из сканированных файлов, преобразованных в формат DjVu,
- программа Poliqarp for DjVu (marasca) (<https://bitbucket.org/jsbien/ndt/wiki/wyniki#!poliqarp-for-djvu-serwer-korpusow-djvu-a-server-for-djvu-corpora>),
- возможность использования регулярных выражений

Словарь как корпус текстов



← → ↺

poliarp.wbl.kf.uw.edu.pl/en/sownik-lindego/

☆ 🔍

Interface language:

English

Change

About

Available texts:

J. Karłowicz, A. Kryński, W. Niedźwiedzki. Dictionary of Polish. Warsaw 1900–1927

S. Bąk, M. R. Mayenowa, F. Peplowski (eds.). Dictionary of the 16th century Polish. Wrocław — Warszawa, 1966-???? (work in progress)

M. Samuel Bogumił Linde. Dictionary of Polish (2nd edition). Lwów 1854-1861.

B. Chlebowski, F. Sulimierski, W. Walewski (ed.), The Geographical Dictionary of the Polish Kingdom and other Slavic Countries, Warszawa 1880-1902

IMPACT GT corpus (1-d), 1570-1756

IMPACT GT corpus (2-d), 1570-1756

Settings

Basic help

A preliminary electronic version of the second edition of Linde's dictionary

Scans were made with Fujitsu fi-6130 scanner and Kofax VRS software by Joanna A. Bilińska. OCR was prepared with FineReader 10 (selected language Polish, default values of other recognition parameters) and saved as PDF/A files with MRC lossy compression. The scans have been converted to DjVu format by Jakub Wilk, he also converted the OCR results to the suitable corpus format.

The corpus consists of ca. 7 million segments. This version of the corpus is available since August 6, 2010.

Query:

Search

It is recommended to switch on graphical concordances option in Settings.

Search can be limited to a specific volume with the `within` clause, e.g. `within vol6 OR within vol6part2`. The clause can be also used to limit the search to the following sections: `body` (entries and their corrections), `errata` (just the corrections), `intro` (introductory texts), `info` (other informations), `varia` (dedications, subscriber lists), `front` (frontmatter), `back` (backmatter).

Структура словарных статей

Структура словарных статей не всегда одинакова. В основном:

- заглавное слово,
- окончания,
- грамматическая информация,
- этимология,
- толкование, дефиниция и синонимы,

Структура словарных статей (2)

- переводы на иностранные языки, начинающиеся с сокращенных названий языков,
- иллюстрации: примеры из литературы со ссылками,
- переносные значения слова и фразеологизмы (также с примерами из литературы со ссылками),
- другие однокоренные слова начинаются с нового абзаца (словообразное гнездо).

Пример словарной статьи в словаре Линде

PIĘĆ, iu; *Boh.* pět; *Slov.* pať; *Sorab.* 2. pēsč; *Sorab.* 1. pietž, piacz, piaczo; *Carn.* pět; *Vind.* pet; *Croat., Bosn. et Rag.* pét; *Slav.* pet; *Eccl. et Ross.* пять; *Graec.* πέντε; *Pers.* peng; *Ind.* pantscha; *Germ.* fünf. *Observes Numeralia cardinalia apud veteres generis feminini*; n. p. One pięć głupich panien wzięwszy lampy, nie wzięły z sobą oleju. 1 *Leop. Math.* 25, 3. On co wziął pięć grzywien, robił niemi; tak iż zyskał niemi drugą pięć. 1 *Leop. Math.* 25, 16. — 2. Gadasz ni w pięć, ni w dziewięć. *Zabl. Amf.* 37. bez sensu, bez składu, nie klei się twa mowa; weder gehauen, noch gestochen. Ożenił się ni w pięć, ni w dziewięć, z kobieta, której nikt nie zna. *Teat.* 36. b, 33. *Mon.* 65, 368. *ib.* 73, 278. *Teat.* 6, 46; *Rag.* ni slastan, ni maastan : ladaco; *Rus.* ni k *selu, ni k *horodu. *Pim. Kam.* 320. (*Slov.* bere pať za dwanast, *deceptor*).
Pochodz. pięćdziesiąt etc., piętnaście etc, piąty, piątek, piątka, popiętny; pięść, piąstka.

Таблица сокращений

- отправная точка – краткий перечень, приведенный,
- данные из части со словами, начинающимися с буквы А и «Правил этимологии» Линде,
- автоматически полученный список слов, написанных курсивом,
- данные, которые случайно встретились

Таблица сокращений (2)

- более 200 сокращений
- около 80 языков и диалектов
- некоторые языки (к примеру, русский) имеют только по одному сокращению
- В большинстве случаев – не менее двух
- к примеру, в случае с венгерским языком – даже 7 сокращений (*Węg.*, *Hung.*, *Hg.*, *Hng.*, *Hungar.*, *Ungar.*, *Ung.*).

Как найти все сокращения для данного языка

- составить запрос по формуле:

$X" \setminus . " \mid Y" \setminus . " \mid Z" \setminus . "$

X, Y, Z – сокращения, $" \setminus . "$ точка,

| - альтернатива

- для венгерского языка это выглядит следующим образом:

$Węg" \setminus . " \mid Hung" \setminus . " \mid Hg" \setminus . " \mid Hng" \setminus . " \mid$
 $Hungar" \setminus . " \mid Ungar" \setminus . " \mid Ung" \setminus . " " . * "$

- 1966-???? (praca w toku)
 - *M. Samuel Bogumił Linde. Słownik języka polskiego (2. wydanie). Lwów 1854-1861.*
 - B. Chlebowski, F. Sulimierski, W. Walewski (red.), *Słownik Geograficzny Królestwa Polskiego i innych krajów słowiańskich*, Warszawa 1880-1902
- Ustawienia
Podstawowa pomoc

2.	Croat. medved, Hung. medve, <i>Etymon.</i> miód, <i>Ross.</i> м	Zakładka
3.	raax, Hung. ros, <i>Vallis.</i> rhyg, <i>Lat. med.</i> r	Zakładka
4.	epa, <i>Graec.</i> ῥαπα , <i>Lat.</i> rapa, Hung. repa, <i>Ital.</i> rapa, rava, <i>Gall.</i> ra	Zakładka
5.	ялокъ, <i>Vind.</i> red, <i>Croat.</i> red, Hung. read, <i>Svec.</i> rad, <i>Lith.</i> redas,	Zakładka
6.	redkou, redkva, <i>Croat.</i> rotkva, Hung. retek, <i>Germ.</i> Rettich, <i>Graecolat.</i>	Zakładka
7.	ъ, <i>Vind.</i> shnura, <i>Croat.</i> snora, Hung. sinor, <i>Germ.</i> Schnur, <i>Niederjäh.</i>	Zakładka
8.	<i>Sorab.</i> eyda, <i>Lapp.</i> atchie, atye, Hung. atya, <i>Turc.</i> ata, <i>Roman.</i> atta,	Zakładka
9.	krzjż, <i>Vind.</i> krish, <i>Croat.</i> kris, Hung. kereszt, <i>Ross.</i> крестъ;	Zakładka
10.	<i>Gr.</i> sakkos, <i>Lat.</i> saccus, Hung. zsak, <i>Lapp.</i> tsakes, <i>Hisp.</i> Gr.	Zakładka
11.	<i>Hisp.</i> Gr. sakkos, <i>Lat.</i> saccus, Hung. zsak, <i>Lapp.</i> tsakes, <i>Hisp.</i>	Zakładka
12.	<i>Svec.</i> aeria, <i>obsol.</i> <i>Gall.</i> error, Hung. Statute vorfommt, <i>Ross.</i> pachat	Zakładka
13.	<i>Svec.</i> aeria, <i>obsol.</i> <i>Gall.</i> error. Hung. aratni, <i>Lat.</i>	Zakładka
14.	um, <i>Lat.</i> <i>Ital.</i> <i>Germ.</i> Dalmaticae et Hung. Venet. 1595. 4. • <i>Etymo</i>	Zakładka
15.	<i>L.</i> Dalmat., Boh., Pol., Ger. et Hung. — studio • <i>Grammatica S</i>	Zakładka
16.	elo Bellosztenēcza, nad to Weg. , Hung. , Hg. , Węgierskie słowa	Zakładka
17.	losztenēcza, nad to Weg. , Hung. , Hg. , Węgierskie słowa te ty	Zakładka
18.	tenēcza, nad to Weg. , Hung. , Hg. , Węgierskie słowa te tylko u	Zakładka
19.	na Teofila. w Wiln. 1594. 4. Weg. Węgierskiego organy. 4.	Zakładka
20.	ledwo <i>Lat.</i> at, atque, et Hung. és, is, <i>Ross.</i> аа; cf. <i>Polon.</i> d	Zakładka
21.	<i>Sorab. inf.</i> - 1 - <i>Dalm.</i> alli; Hung. avvagy; <i>Bosn.</i> il, illi, alli; <i>Ra</i>	Zakładka
22.	<i>Rab.</i> damascak • <i>Croat.</i> et Hung. damask • <i>Vind.</i> tomashk •	Zakładka

Результат поиска всех сокращений для венгерского языка

Węg"\" | Hung"\" | Hg"\" | Hng"\" | Hungar"\" | Ungar"\" | Ung"\" \".*"

Wyswietlanie wynikow 1-25

1.	żnyi, u Rosssyan pżekiił, pżekozł znaczy rażać	Zakładka
2.	wybi ^{er} am, wiłł, wabił; u Rosssyan woliu znaczy chce, ży- bej	Zakładka
3.	ieda chł ^o dnik, kuczka letnia; u Rosssyan besieda Mundarten richten. Bi	Zakładka
4.	Chytry u Rosssyan i Czechów toż samo co u	Zakładka
5.	Pan; u Rosssyan gospod (cf. gospodarz); u Sor	Zakładka
6.	y u Czechów to co i u nas; u Rosssyan zaś upria- dieß lehte bedentei	Zakładka
7.	Narzędzie u nas instrument; u Rosssyan Na- den Polen serce; bejm	Zakładka
8.	ment lub naczynie nadoba; u Rosssyan nadobe Wenden kniez (der Zi	Zakładka
9.	m przywarę, i t. d. Gramota u Rosssyan dy- sterfrau, die Ronne,	Zakładka
10.	bey den Croaten zählen, rechnen, und broi die Rabi, die	Zakładka
11.	ryków szczególnie imieniny, u Rosssyan god Günstigkeit, pogoda günsf	Zakładka
12.	o się, nie tylko u nas, lecz i u Russen god Jahr, godina Zeit, Schid	Zakładka
13.	uo zizowia juc, imuc. — Głowiańskie i	Zakładka

Результаты поиска для фраз типа «у россиян»

$$u \cdot n$$

Wyniki

Znaleziono 475 wyników

Wyświetlanie wyników 1-25

- | | | |
|-----|---|--------------------------|
| 1. | też u współpracowników Lindego, którzy | Zakładka |
| 2. | en czas praca była wielka u Franciszkanów (od których czegoś jenc | Zakładka |
| 3. | miennie nabył) i u Bożociałków, z którymi jeszcze nie m | Zakładka |
| 4. | wie u Norbertanów, zwłaszcza co do Fragment | Zakładka |
| 5. | ne (zład u Windów <i>dobnik zegar</i>): u Czechów | Zakładka |
| 6. | ad u Windów <i>dobnik zegar</i>): u Czechów odrócz czasu. punktu czasu | Zakładka |
| 7. | Biel. Kr. 53. cf. ład, ładny; u Czechów zaś Ladon, Smrto- | Zakładka |
| 8. | nos, u Krainczyków Ladon, Tor, Tòrk <i>Mars</i> , | Zakładka |
| 9. | Mercury. Cerera, Marzanka, u Czechów Zewena, u Kra- | Zakładka |
| 10. | va toż co Dvana. Bachus u Krainczyków Post. | Zakładka |

Результаты поиска для фраз типа «у чехов»

u .*ów

Wyniki

Znaleziono 554 wyniki

Wyświetlanie wyników 176-200

Poprzednich 25

176.	v prawie też sa- wie się po Niemiecku Stüdfen, co się znaczy pi	Zakładka
177.	albo <i>intestinum coecum</i> , po Polsku kątlica. Urzęd. 30. —	Zakładka
178.	Post. W. 2, 77. 'Katolicko, po katolicku adv., TOWCZYK, a. m.	Zakładka
179.	n. p. Po katolicku mówiąc. Torz. 276. 'Niekatol	Zakładka
180.	z. f. (t. j. pozabijał ich). Po katowsku, *katowsko	Zakładka
181.	zaczął zwadę, Ale po kawalersku, ale nie przez zdradę. KA	Zakładka
182.	rdz. Trag. 197. (meźnie). Po kawalersku wytnę się Kawecan dl	Zakładka
183.	ją letko po kawalersku. Zabł. Roz. 119. (z pańska)	Zakładka
184.	władzą królewską zagajony, co po Polsku snadniej	Zakładka
185.	mi widzę zaczynasz prawić po kaznodziejsku. Teat. 52. d, KCIEĆ, kei	Zakładka
186.	pozwo- w człowieku <i>ilium</i> , po Polsku kielbaśnica. Urzęd. 30. ber	Zakładka
187.	mek i miasto w Kurlandyi, po Polsku Kiesią nazywają.	Zakładka

Результаты поиска для фраз типа «по-польски»

(po|Po)(.*sku|.*cku))

Wyświetlanie wyników 26-50

Poprzednich 25

26.	uszczoneym, ПОПУСТИТЕЛЬНЫМ, w Polskim zaś dopuszczalny zna-	Zakładka
27.	zyku, tak i w Polskim, niemasz powszechnej zgody	Zakładka
28.	ów. Co do Słowiańszczyzny, w Czeskim i Bośnić- języku baltic. —	Zakładka
29.	t, hrst, prs, prsk, i t. d. a w Kraińskim lauter geschrieben, z. B. smi	Zakładka
30.	2) Które głoski w Słowiańskim stoją za Łacińsko-Teu-	Zakładka
31.	min, piszą, nawet w Polskim, hruby, hrubo i t. d. zamiast	Zakładka
32.	kim abecadla wcale nie było; w Polskim musi być zachó- Alphabet	Zakładka
33.	5) Słowa w Greckim języku a spiritu aspero (h) z:	Zakładka
34.	w drugim przypadku doczery. Croat. i	Zakładka
35.	za nieczemne zaś uchodzi w drugim; w jednym znaczy po- len a	Zakładka
36.	chwałę, w drugim naganę lub wzgardę; w jed	Zakładka
37.	zaletę, w drugim przywarę, i t. d. Gramota u	Zakładka
38.	re przecież choć w dalekim stosunku się schodzą, służy c	Zakładka

Результаты поиска для фраз типа «в венгерском»
w.*im

Zapytanie: z .iego

Szukaj

Wyniki

Znaleziono 585 wyników

Wyświetlanie wyników 26-50

Poprzednich 25

26.	sto nie wszczyna się z drugiego ramie, i nie są w obudwóch	Zakładka
27.	ten kształt i u nas śpital, z Łacińskiego Hospi- rib Pol. grzyb; no	Zakładka
28.	dokładniej wyprowadzać, jak z Rossyjskiego Korowod, choro- słoune m	Zakładka
29.	kiego Śchlacht bitwa; lecz z Niemieckiego Gefchlecht, wy- (man ve	Zakładka
30.	; że zaś tkwi w pochodzących z niego, Jahr verlieben. Erst wenn ma	Zakładka
31.	wyrażnie, istotne głoski źródłosłowa swego zachow	Zakładka
32.	pochodzących. Tak z Greckiego, stan, stenai (Lat. stare,	Zakładka

Результаты поиска для фраз типа «с греческого»

z .*iego

Проблемы

- Неспециалист может испытывать определенные трудности в работе с системой,
- включить непосредственно в корпус таких сведений, как развитие сокращений,
- разработка возможности появления в тексте словаря развития аббревиатур при подведении курсора

:	zad
<i>n sąsiedztwie, obok czegoś; w</i>	wy
<i>ty stykają się ze sobą lub jest</i>	wo
<i>Miech, Vulg, PolAnt, Mącz,</i>	wic
<i>ch, Vulg, PolAnt, JanStat; se-</i>	spo
<i>cz; circa Miech, Vulg, Mącz,</i>	iać
<i>per Vulg, PolAnt. prone Vulg</i>	kor
<i>Mącz, Cn; coram, supra i or</i>	ni,
	Ko
<i>ch sie (lub odv ruch iest nieis-</i>	nał

Вид развития для аббревиатуры

(фрагмент «Словаря польского языка XVI века»)

Для кого словарь?

- может использоваться исследователями, обрабатывающими исторические словари с целью сверки сносок к словарю С.Б.Линде
- интересует специалистов-филологов, лиц, изучающих лексику, историю и культуру этого периода, а также лексикографов
- может использоваться не только филологами-полонистами, но и другими славистами

Литература

- Janusz S. Bień, *Efficient search in hidden text of large DjVu documents*. B: *Advanced Language Technologies for Digital Libraries. Lecture Notes in Computer Science (Theoretical Computer Science and General Issues)* (6699), 2011. Springer, стр. 1-14, <http://bc.klf.uw.edu.pl/177/>.
- Janusz S. Bień, *Scanned publications in digital libraries: new Open Source DjVu tools*. B: *the Library 2.012 worldwide virtual conference*, October 3 - 5, 2012, Internet, <http://bc.klf.uw.edu.pl/298/>

- Samuel Bogumił Linde, *Słownik języka polskiego*, изд. I, Warszawa 1807-1814,
<http://kpbc.umk.pl/publication/8173>.
- Samuel Bogumił Linde, *Słownik języka polskiego*, изд. II, Zakład Narodowy im. Ossolińskich, Lwów 1854-1860, <http://poliqarp.wbl.klf.uw.edu.pl/slownik-lindego/>.
- Samuel Bogumił Linde, *Słownik języka polskiego*, изд. II, репринт, Wydawnictwo Gutenberg-Print, Warszawa 1994,
<http://poliqarp.wbl.klf.uw.edu.pl/extra/linde/index.djvu>.

- Samuel Bogumił Linde, *Słownik języka polskiego*, оцифрованная версия с поисковой системой, <http://poliarp.wbl.klf.uw.edu.pl/extra/linde/index.djvu>.
- Bożena Matuszczyk, *Słownik języka polskiego S. B. Lindego. Warsztat leksykografa*, Wydawnictwo KUL, Lublin 2006.
- *Słownik polszczyzny XVI wieku*, Polska Akademia Nauk. Instytut Badań Literackich, Zakład Narodowy im. Ossolińskich, Wrocław, 1966-, <http://www.spxvi.edu.pl/spxvi/slownik.php?s=2>.