

XXVII Szkoła Matematyki

Poglądowej

Gołąb et al. 1970

język

język polski

Rada Języka Polskiego

„kulturyści”

Gołąb et al. 1970

NEP PWN t. 6, 1996

WEP PWN t. 5, 1965

Konkordancje

Teksty

Teksty w poligrafii

Teksty w informatyce

Teksty w informatyce

Desktop publishing

Terminologia

Korpusy

Korpus słownika frekwencyjnego

Korpusy języka polskiego

Korpusy języka polskiego

Metodologia

Przykładowe wyniki

Analizator morfologiczny SAM

Analizator morfologiczny SAM

Zastosowania analizatora SAM

Zastosowania analizatora SAM

Zastosowania analizatora SAM

Przykładowe wyniki

Zastosowania gramatyki
Szpakowicza

Przykładowe wyniki

Analizatory syntaktyczne
AMOS i AS

Analizator syntaktyczny AMOS

Projekt KBN nr 8 S503 032 27

analizator syntaktyczny AS

analizator syntaktyczny AS

Projekt KBN nr 8 T11C 002 13

Korpusy

Wyniki

Metodologia

*Matematyka w informatyce
i vice versa*

24–28 sierpnia 2001

Janusz S. Bień

Język naturalny i komputery

(komputery i teksty)

25 sierpnia 2001

Uniwersytet Warszawski

Instytut Informatyki

dr hab. JANUSZ S. BIEŃ, prof. UW

(od 1 lipca 1998)

Zakład Zastosowań Informatycznych

Instytut Orientalistyczny

Ferdynand de Saussure
(1857-1913)

language

mowa

parole

mówienie

langue

język

Zbigniew Gołąb, Adam Heinz, Kazimierz Polański

Słownik terminologii językoznawczej

Państwowe Wydawnictwo Naukowe

Warszawa 1970

MOWA

Zjawisko wzajemnego porozumiewania się ludzi głównie za pomocą dźwięków

MÓWIENIE

Akt indywidualny konkretny polegający w zasadzie na kojarzeniu przez osobę mówiącą pewnej treści pojęciowej z formą utworzoną na mocy określonego przedstawienia dźwiękowego . . .

JĘZYK

Zespół społecznie wytworzonych i obowiązujących znaków dźwiękowych (względnie wtórnie pisanych) oraz reguł określających ich użycie, a funkcjonujących jako narzędzie komunikacji społecznej.

DIALEKT

Odmiana języka ogólnonarodowego wyróżniająca się pewnymi cechami fonetycznymi lub fonologicznymi i leksykalnymi, rzadziej morfologicznymi i składniowymi, używana na określonym terytorium i przez określoną warstwę społeczną ...

NARZECZE, GWARA, ...

JEZYK

Dialekt z własną flotą i armią

<http://www.ethnologue.com/>

Barbara F. Grimes (ed.)

Ethnologue: Languages of the World

14th Edition.

SIL International 2000

Total Living Languages: 6809

<http://www.sil.org/ethnologue/top100.html>

- 1 CHINESE, MANDARIN [CHN] China 885,000,000
- 2 ENGLISH [ENG] United Kingdom 322,000,000
- 3 SPANISH [SPN] Spain 266,000,000
- ...
- 23 GUJARATI [GJR] India 44,000,000

- 24 POLISH [PQL] Poland 44,000,000

- 25 ARABIC, EGYPTIAN SPOKEN [ARZ] Egypt 42,500,000
- 26 UKRAINIAN [UKR] Ukraine 41,000,000
- 27 ITALIAN [ITN] Italy 40,000,000
- ...

Jednym z tych trzech czy pięciu tysięcy języków i ich odmian mających zemrzeć przed końcem XXII w. może być język polski.

Prof. dr hab. **Walery Pisarek**

Przewodniczący Rady Języka Polskiego
przy Prezydium Polskiej Akademii Nauk

Walery Pisarek, Jolanta Rokoszowa

Prawne ramy troski o język

Jan Miodek, red. *O zagrożeniach i bogactwie polszczyzny*,

Wrocław 1996, s. 47–67

Kto zagwarantuje, że po wejściu do Unii Europejskiej język polski nie stanie się jednym z dialektów lokalnych, takim jak łużycki czy kaszubski?

Andrzej IBIS Wróblewski

członek Rady Języka Polskiego

przy Prezydium Polskiej Akademii Nauk

Rzeczpospolita nr 174 (5644) z 27 lipca 2000 r.

Niektórzy utrzymują, że mowa ojczysta przestanie istnieć w roku 2008

Prof. dr hab. Jan Miodek

członek Rady Języka Polskiego

przy Prezydium Polskiej Akademii Nauk

Jan Bończa-Szabłowski

Wywiad z członkami RJP

(prof. Jadwigą Puzyniną, prof. Jerzym Bralczykiem i prof. Janem Miodkiem)

Rzeczpospolita nr 88 (4948) z 15 kwietnia 1998 r.

<http://www.rjp.pl/opinie.html>

Słowo *internet* po jakimś czasie zaczęło nazywać każdą sieć informatyczną ...

<http://www.rjp.pl/opinie.html>

Słowo *internet* po jakimś czasie zaczęło nazywać każdą sieć informatyczną — w tym znaczeniu jest pisane małą literą.

Redakcja naukowa

prof. Andrzej Markowski

**Wielki słownik ortograficzny
języka polskiego**

Wilga, Warszawa 1999

s. 177

Internet (nazwa własna systemu)

internet (każdy system łączności komputerowej)

Redakcja naukowa

prof. Andrzej Markowski

Ortograficzny słownik języka polskiego

Wilga, Warszawa 2000

s. 301

Internet (nazwa własna systemu)

internet (każdy system łączności komputerowej)

prof. dr hab. *Andrzej Markowski*

wiceprzewodniczący Rady Języka Polskiego
dyrektor Instytutu Języka Polskiego UW

...

Pamięć maszyn matematycznych jest tak skonstruowana, że najpierw są usytuowane pojęcia o szerszym zakresie, a potem o zakresach węższych.

Otóż wprowadzenie nowego systemu oznaczania dat jest uzasadnione tym, że we współpracy międzynarodowej zarówno gospodarczej, jak i naukowo-technicznej coraz szerzej są wykorzystywane maszyny matematyczne. Pamięć maszyn matematycznych jest tak skonstruowana, że najpierw są usytuowane pojęcia o szerszym zakresie, a potem o zakresach węższych. Z tej właśnie zasady wynika kolejność: rok, miesiąc, dzień, a nie dzień, miesiąc, rok. Powszechnie wprowadzenie jednolitego systemu ułatwia oficjalne kontakty międzynarodowe oraz umożliwia wykorzystanie środków automatycznego przetwarzania danych.

M.S. (Mieczysław Szymczak)

Objaśnienia wyrazów i zwrotów

O nowym sposobie numerycznego oznaczania dat
(tekst audycji Radiowego Poradnika Językowego)

Poradnik Językowy nr 8 (427), s. 545–548.

Pismo, które powinno się znaleźć w każdej bibliotece szkolnej
(komunikat nr 142/PR4-552-29/68 ogłoszony w Dzienniku
Urzędowym MOiSW z dn. 31 XII 1968 r.)

Redaktor naukowy
prof. dr Mieczysław Szymczak

*Słownik ortograficzny języka polskiego
wraz z zasadami pisowni i interpunkcji*

Państwowe Wydawnictwo Naukowe
Warszawa

prof. dr hab. *Mieczysław Szymczak*

członek rzeczywisty PAN

przewodniczący Komitetu Językoznawstwa PAN

dyrektor Instytutu Języka Polskiego UW

...

TEKST

[...] utrwalone w pamięci lub na piśmie
wypowiedzenie lub ich szereg ...

Ciąg linearny elementów i struktur językowych
(fonicznych, morfologicznych, syntaktycznych)
tworzony określoną strukturę relacyjną
(pozycyjną).

łc. *textum* tkanina, związek (słów)

TEKSTOLOGIA

dziedzina wiedzy o literaturze, zajmująca się krytycznym opracowaniem tekstu — ustaleniem jego pochodzenia (autorstwa, czasu powstania) oraz poprawności, czyli zgodności kształtu zapisu grafemicznego) językowego z intencją autora

KONKORDANCJA

rel. alfabetyczny spis wyrazów lub zdań, znajdujących się w *Biblii*, z dokładnym uwzględnieniem ich miejsca;

k. wyrazowa zestawia poszczególne słowa (pierwszą taką k. sporządził Hugon z Sancto Caro, zm. 1263) ...

tego rodzaju k. sporządzone są także dla wybitnych dzieł literatury pięknej.

		pies	
D0325	a dzieci pluskały się w wodzie. Pies	SSNA-----P	był nieodłącznym towarzyszem
E0452	choć chodziłam za nim jak ten pies.	SSNA-----P	Tyś przy nim tak nie warowała,
E0885	do nogi! Bobik, leżeć! Zdechł pies!	SSNA-----P	
E1221	żeby bez plakatów, bo i tak pies	SSNA-----P	z kulawą nogą nie przyjdzie.
E1245	Co to? .. Pies ..	SSNA-----P	Pies? Jaki pies? Gryzie? ..
E1245	Co to? .. Pies .. Pies?	SSNA-----P	Jaki pies? Gryzie? .. Czasami
E1245	Co to? .. Pies .. Pies? Jaki pies?	SSNA-----P	Gryzie? .. Czasami .. On się
E1245	.. On się ciebie słucha, ten pies?	SSNA-----P	To każ mu wyjść spod ławki
C0790	Jeśli gwizdniemy na wyszkolonego psa	SSAA-----P	specjalnym gwizdkiem
C1207	psu dodatkowe serce (z innego psa)	SSGA-----P	tak, że ten żył z dwoma
D0881	w szarówce ledwie ustępującej nocy psa	SSAA-----P	Kolędy biegnącego ze skowytem
D1207	a kochałem tylko Grażynę i psa.	SSAA-----P	Układaliśmy stóg. Najpierw na
E1017	fotografowali. Domy, cegielnię, psa	SSAA-----P	przy budzie, uschnięte drzewo
C1207	skupił on uwagę i wysiłki na psach,	SPLA-----P	przeszczepiając im płuca,
D1005	Pachołka, córce zaszczutego psami	SPIA-----P	Pachołka? Że staraj się, aby
D1333	Poszliśmy we trzech, z psem,	SSIA-----P	na nadbrzeże, przy którym
E0996	Kondycja fizyczna pod psem.	SSIA-----P	Dwadzieścia lat klęczałem w
B1285	oraz schronisko dla bezdomnych psów,	SPGA-----P	kotów i tym podobne.
D1221	Leosi. Już z daleka szczekanie psów	SPGA-----P	przywodziło na myśl rozległe
E1251	Nie. Po prostu miałem już tyle psów,	SPGA-----P	że naprawdę znam się z
E1251	jak w leśniczówce, zawsze było masę psów.	SPGA-----P	Nawet w czasie wojny trzymałem
B0835	zwierzę. Tylko Ewa ruszyła psu	SSDA-----P	z pomocą. Wzięła go na rękę
C1207	doktór Demichow wszczepił psu	SSDA-----P	dodatkowe serce (z innego psa)
D0492	guzik. Innymi słowy, panie, to dwa psy,	SPNA-----P	które usiłują jeden drugiemu
D0640	Biały promień szastnął w krzaki, psy	SPNA-----P	z ujadaniem skoczyły ku
D1461	w zapadłej tylko co ciemności. Psy	SPNA-----P	nagonki, które od pewnej

mechanizacja i komputeryzacja

obliczenia

literackie i lingwistyczne

literary and linguistic computing

ASSOCIATION OF LLC

Journal of ALLC, ALLC Conferences . . .

Jadwiga Sambor

Słowa i liczby

Zagadnienia językoznawstwa statystycznego

Ossolineum:

Wrocław 1972

Adam Pawłowski

Séries temporelle en linguistique.

Avec application à l'attribution de textes:

Roman Gary et Emile Ajar

Champion: Paris 1998

Adam Pawłowski

Metody kwantytatywne

w sekwencyjnej analizie tekstu

Uniwersytet Warszawski, Katedra Lingwistyki Formalnej:

Warszawa 2001

mechanizacja i komputeryzacja składu
skład gorący

- linotypy i TTS (*TeleTypesetting*)
- tastry i monotypy

mechanizacja i komputeryzacja składu
skład zimny

- taśma perforowana TTS
- dyskietki *hard sector*
- taśmy magnetyczne
- opogramowanie jako składnik sprzętu

- teksty programów
- teksty komentarzy
- teksty dokumentacji
- teksty publikacji

Informatyczne programy składu

- roff (*runoff*)
- nroff (*new roff*)
- ...
- T_EX, L^AT_EX itd.

osobista drukarka laserowa

PostScript

profesjonalne naświetlarki

Korpus

David Crystal

A Dictionary of Linguistics and Phonetics,
Blackwell, 3rd Edition, 1991.:

kolekcja danych lingwistycznych — tekstów pisanych lub transkrypcji wypowiedzi mówionych — która może stanowić punkt wyjścia do opisu lingwistycznego lub do weryfikacji pewnych hipotez dotyczących języka

Korpus

John Sinclair

Corpus, Concordance, Collocation,
Oxford University Press 1991.:

kolekcja naturalnie występujących tekstów języka naturalnego, dobranych w celu scharakteryzowania aktualnego stanu języka naturalnego lub jego różnorodności

I. Kurcz, A. Lewicki, J. Sambor,
K. Szafran, J. Woronczak,
*Słownik frekwencyjny współczesnej polszczyzny
pisanej*
Instytut Języka Polskiego PAN
Kraków 1990.

opracowanie tomu zbiorczego i skład komputerowy:
Krzysztof Szafran
Instytut Informatyki UW

		pies	
D0325	a dzieci pluskały się w wodzie. Pies	SSNA-----P	był nieodłącznym towarzyszem
E0452	choć chodziłam za nim jak ten pies.	SSNA-----P	Tyś przy nim tak nie warowała,
E0885	do nogi! Bobik, leżeć! Zdechł pies!	SSNA-----P	
E1221	żeby bez plakatów, bo i tak pies	SSNA-----P	z kulawą nogą nie przyjdzie.
E1245	Co to? .. Pies ..	SSNA-----P	Pies? Jaki pies? Gryzie? ..
E1245	Co to? .. Pies .. Pies?	SSNA-----P	Jaki pies? Gryzie? .. Czasami
E1245	Co to? .. Pies .. Pies? Jaki pies?	SSNA-----P	Gryzie? .. Czasami .. On się
E1245	.. On się ciebie słucha, ten pies?	SSNA-----P	To każ mu wyjść spod ławki
C0790	Jeśli gwizdniemy na wyszkolonego psa	SSAA-----P	specjalnym gwizdkiem
C1207	psu dodatkowe serce (z innego psa)	SSGA-----P	tak, że ten żył z dwoma
D0881	w szarówce ledwie ustępującej nocy psa	SSAA-----P	Kolędy biegnącego ze skowytem
D1207	a kochałem tylko Grażynę i psa.	SSAA-----P	Układaliśmy stóg. Najpierw na
E1017	fotografowali. Domy, cegielnię, psa	SSAA-----P	przy budzie, uschnięte drzewo
C1207	skupił on uwagę i wysiłki na psach,	SPLA-----P	przywodziło na myśl rozległe
D1005	Pachołka, córce zaszczutego psami	SPIA-----P	Pachołka? Że staraj się, aby
D1333	Poszliśmy we trzech, z psem,	SSIA-----P	na nadbrzeże, przy którym
E0996	Kondycja fizyczna pod psem.	SSIA-----P	Dwadzieścia lat klęczałem w
B1285	oraz schronisko dla bezdomnych psów,	SPGA-----P	kotów i tym podobne.
D1221	Leosi. Już z daleka szczekanie psów	SPGA-----P	przywodziło na myśl rozległe
E1251	Nie. Po prostu miałem już tyle psów,	SPGA-----P	że naprawdę znam się z
E1251	jak w leśniczówce, zawsze było masę psów.	SPGA-----P	Nawet w czasie wojny trzymałem
B0835	zwierzę. Tylko Ewa ruszyła psu	SSDA-----P	z pomocą. Wzięła go na rękę
C1207	doktór Demichow wszczepił psu	SSDA-----P	dodatkowe serce (z innego psa)
D0492	guzik. Innymi słowy, panie, to dwa psy,	SPNA-----P	które usiłują jeden drugiemu
D0640	Biały promień szastnął w krzaki, psy	SPNA-----P	z ujadaniem skoczyły ku
D1461	w zapadłej tylko co ciemności. Psy	SPNA-----P	nagonki, które od pewnej

Marek Łaziński

Textkorpora in Polen. Ein Überblick.

Katharina Böttger, Sabine Dönninghaus, Robert Marzari (Hrsg.)

*Beiträge der Europäischen Slavistischen Linguistik
(POLYSLAV)*, Band 4.

Peter Rehder, Igor Smirnov (Hrsg.)

Die Welt der Slaven. Sammelbände

Band 12.

Verlag Otto Sagner, München 2001,

pp 123-133.

KORPUSY W LEKSYKOLOGRAFII

Wydawnictwo Naukowe PWN

Redakcja Słowników Języka Polskiego

<http://slovniki.pwn.pl/korpus/>

Mirosław Bańko (red.)

Inny słownik języka polskiego

obliczenia

literackie i lingwistyczne

literary and linguistic computing

działalność usługowa

wobec tradycyjnych działów nauki

maszynowe tłumaczenie

Machine Translation

lingwistyka komputerowa

Computational Linguistics

sztuczna inteligencja

Artificial Intelligence

lingwistyka informatyczna
(Computational Linguistics)

badanie języka naturalnego z punktu widzenia
potrzeb i możliwości przetwarzania tekstów

Prof. Witold Doroszewski

Elementy leksykologii i semiotyki

PWN: Warszawa 1970, s. 242

Antoine Meillet (1866–1936)

Wyrazu zasadniczo zdefiniować nie można

SYSTEM ANALIZY MORFOLOGICZNEJ

ftp://ftp.mimuw.edu.pl/pub
/users/polszczyzna/SAM-95/

Krzysztof Szafran

Analizator morfologiczny SAM-95

— *opis użytkowy.*

Raport Instytutu Informatyki Uniwersytetu Warszawskiego
TR 96-05 (226), maj 1996

SYSTEM ANALIZY MORFOLOGICZNEJ

`ftp://ftp.mimuw.edu.pl/pub`
`/users/polszczyzna/SAM-95/`

jeden z wielu, ale **jedyny**
analizator morfologiczny języka polskiego
dostępny bezpłatnie w Internecie

Erjavec, T., Lawson, A. (eds).

East meets West

— *A Compendium of Multilingual Resources*. CD-ROM.

TELRI Association [1998].

Trans-European

Language Resources Infrastructure

COPERNICUS Concerted Action

PODSTAWA TEORETYCZNA:

Jan Tokarski

**Schematyczny indeks *a tergo*
polskich form wyrazowych**

Opracowanie i redakcja:

Zygmunt Saloni

Wydawnictwo Naukowe PWN

Warszawa 1993.

K.Szafran: *Automatyczne hasłowanie tekstu polskiego*, **POLONICA XVIII**, s. 51–64, Kraków 1997.

K.Szafran: *Automatic Lemmatisation of Texts in Polish — Is It Possible?*. In: Uwe Junghans, Gerhild Zybatow (eds.): **Formale Slavistik**, pp. 437–442, Vervuert Verlag, Frankfurt am Main, 1997.

K.Szafran: SAM-96 — The Morfological Analyser for Polish, In: A.S. Narin'yani (ed.): **Proceedings of International Workshop DIALOGUE'97 Computational Linguistics and its Applications**, Yasnaya Polyana, Russia, June, 10-15, 1997, pp. 304–308, Moskwa, 1997.

J.S.Bień, K.Szafran: Tokarski's Morphological Analysers of Polish, In: R.Marcinkeviciene, N.Volz (eds.): **Proceedings of the Second European Seminar Language Applications for a Multilingual Europe**, Kaunas, Lithuania, April 17-20, 1987, pp. 11-16, Mannheim-Kaunas, 1987.

Uniwersytet Adama Mickiewicza w Poznaniu
Wydział Matematyki i Informatyki
projekt POLENG

KBN 8 T11C 011 16

Komputerowe tłumaczenie z języka polskiego na angielski
tekstów informatycznych umieszczonych na stronach WWW

<http://poleng.wmid.amu.edu.pl>

F. Graliński, Hasłowanie korpusu polskich tekstów
informatycznych (1.2 mln słów), *Technologia mowy i języka*
t. 4, Poznań 2000, s. 147-153.

Instytut Informatyki Politechniki Śląskiej

BW 440/RAU2/97

KBN 8 T11C 007 17

(Translacja tekstów w języku polskim na język migowy)

P. Fabian, A. Migas, N. Suszczańska, Zastosowania analizy morfologicznej i składniowej w procesie rozpoznawania mowy. *Technologia mowy i języka* t. 3, Poznań 1999, s. 155-165.

N. Suszczańska, M. Forczek, A. Migas, Wieloetapowy analizator morfologiczny, *Technologia mowy i języka* t. 4, Poznań 2000, s. 155-165.

Instytut Matematyki Uniwersytetu Gdańskiego

korektor ortograficzny MSPELL

`ftp://ftp.gust.org/mspell-1.0.tar.gz(???)`

Włodek Bzyl, Detection and correction of spelling errors in marked-up documents, *Paperless T_EX. EuroT_EX 99 Proceedings*, pp 290–307.

		pies	
D0325	a dzieci pluskały się w wodzie. Pies	SSNA-----P	był nieodłącznym towarzyszem
E0452	choć chodziłam za nim jak ten pies.	SSNA-----P	Tyś przy nim tak nie warowała,
E0885	do nogi! Bobik, leżeć! Zdechł pies!	SSNA-----P	
E1221	żeby bez plakatów, bo i tak pies	SSNA-----P	z kulawą nogą nie przyjdzie.
E1245	Co to? .. Pies ..	SSNA-----P	Pies? Jaki pies? Gryzie? ..
E1245	Co to? .. Pies .. Pies?	SSNA-----P	Jaki pies? Gryzie? .. Czasami
E1245	Co to? .. Pies .. Pies? Jaki pies?	SSNA-----P	Gryzie? .. Czasami .. On się
E1245	.. On się ciebie słucha, ten pies?	SSNA-----P	To każ mu wyjść spod ławki
C0790	Jeśli gwizdniemy na wyszkolonego psa	SSAA-----P	specjalnym gwizdkiem
C1207	psu dodatkowe serce (z innego psa)	SSGA-----P	tak, że ten żył z dwoma
D0881	w szarówce ledwie ustępującej nocy psa	SSAA-----P	Kolędy biegnącego ze skowytom
D1207	a kochałem tylko Grażynę i psa.	SSAA-----P	Układaliśmy stóg. Najpierw na
E1017	fotografowali. Domy, cegielnię, psa	SSAA-----P	przy budzie, uschnięte drzewo
C1207	skupił on uwagę i wysiłki na psach,	SPLA-----P	przeszczepiając im płuca,
D1005	Pachołka, córce zaszczonego psami	SPIA-----P	Pachołka? Że staraj się, aby
D1333	Poszliśmy we trzech, z psem,	SSIA-----P	na nadbrzeże, przy którym
E0996	Kondycja fizyczna pod psem.	SSIA-----P	Dwadzieścia lat klęczałem w
B1285	oraz schronisko dla bezdomnych psów,	SPGA-----P	kotów i tym podobne.
D1221	Leosi. Już z daleka szczekanie psów	SPGA-----P	przywodziło na myśl rozległe
E1251	Nie. Po prostu miałem już tyle psów,	SPGA-----P	że naprawdę znam się z
E1251	jak w leśniczówce, zawsze było masę psów.	SPGA-----P	Nawet w czasie wojny trzymałem
B0835	zwierzę. Tylko Ewa ruszyła psu	SSDA-----P	z pomocą. Wzięła go na rękę
C1207	doktór Demichow wszczepił psu	SSDA-----P	dodatkowe serce (z innego psa)
D0492	guzik. Innymi słowy, panie, to dwa psy,	SPNA-----P	które usiłują jeden drugiemu
D0640	Biały promień szastnął w krzaki, psy	SPNA-----P	z ujadaniem skoczyły ku
D1461	w zapadłej tylko co ciemności. Psy	SPNA-----P	nagonki, które od pewnej

Stanisław Szpakowicz

Automatyczna analiza składniowa
polskich zdań pisanych.

Praca doktorska (promotor S. Waligórski),
Instytut Informatyki UW 1978.

Formalny opis składniowy zdań polskich, 2. wyd.
Wydawnictwa Uniwersytetu Warszawskiego 1986.

`ftp://ftp.mimuw.edu.pl/pub`
`/users/polszczyzna/Szpakowicz/`

E. Kubacka, P. Makowski

Generowanie analizatorów składniowych
na podstawie gramatyki
wybranego podzbioru języka polskiego
Prace IPI PAN nr 700, Warszawa 1991

E. Dobryjanowicz

Podstawy przetwarzania języka naturalnego
Wybrane metody analizy składniowej
Seria *Problemy Współczesnej Nauki i Techniki*
Akademicka Oficyna Wydawnicza RM: Warszawa 1992

Uniwersytet Adama Mickiewicza w Poznaniu
Wydział Matematyki i Informatyki
projekt POLENG

KBN 8 T11C 011 16

Komputerowe tłumaczenie z języka polskiego na angielski
tekstów informatycznych umieszczonych na stronach WWW

<http://poleng.wmid.amu.edu.pl>

K. Jassem, Dealing with Free Order and Non-Language
Markers in a Top-Down Left-First Algorithm, *Technologia
mowy i języka* t. 4, Poznań 2000, s. 147-153.

LINGWISTYCZNA PODSTAWA TEORETYCZNA:

Marek Świdziński

Formalny opis polskich zdań o składniku zdaniowym.

Praca habilitacyjna. Maszynopis powielony, Instytut Języka Polskiego UW, Warszawa 1987.

Gramatyka formalna języka polskiego.

Wydawnictwa Uniwersytetu Warszawskiego 1992.

INFORMATYCZNA PODSTAWA TEORETYCZNA:

Gramatyki metamorficzne Colmerauera

Bień, J. S.

Processing Polish with metamorphosis grammars.

Tübingen Workshop on Computational
Linguistics,

Tübingen, 15–28 September 1996,

[ftp://ftp.mimuw.edu.pl/pub/users/jsbien/
tybinga96/pol-mg.ps](ftp://ftp.mimuw.edu.pl/pub/users/jsbien/tybinga96/pol-mg.ps)

Janusz S. Bień

Komputerowa weryfikacja
formalnej gramatyki Świdzińskiego.

Biuletyn Polskiego Towarzystwa Językoznawczego,
zeszyt LII (1996), s. 147–164.

Komputerowa weryfikacja opisu składni polskiej.
Raport Instytutu Informatyki Uniwersytetu Warszawskiego
TR 96–06 (227), maj 1996.

[ftp://ftp.mimuw.edu.pl/pub/users/
polszczyzna/AMOS-95/](ftp://ftp.mimuw.edu.pl/pub/users/polszczyzna/AMOS-95/)

**Analizator
morfologiczno-syntaktyczny
dla obszernego podzbioru
języka polskiego**

kierownik projektu
Janusz S. Bień

1994 — 1996

AS:

język Prolog

obszerne fragmenty kodu AMOSa

wstępna transformacja gramatyki

analiza wstępująca

zapamiętywanie wyników częściowych

„świdzińskizacja” rezultatów

AS:

przetworzono około 2000 przykładów
utworzono ponad 100 000 drzew rozbioru
gramatycznego
typowe zdanie przetwarzane 1 sekundę
typowe zdanie ma 10 różnych interpretacji
syntaktycznych

Zestaw testów do weryfikacji i oceny analyzerów języka polskiego

kierownik projektu
Janusz S. Bień

1 lipca 1997 — 31 grudnia 1999

treebanks

bank rozbiorów gramatycznych

Płyta CD

Lingwistyczne zasoby polszczyzny

Pełne sprawozdanie merytoryczne
i przykładowe pliki:

```
ftp://ftp.mimuw.edu.pl/pub  
/users/polszczyzna/tajp/
```

inżynieria językowa

language engineering

Continuous speech recognition
ARPA 1971-1977

HARPY
Carnegie-Mellon University

dr hab. JANUSZ S. BIEŃ, prof. UW

jsbien@mail.uw.edu.pl

<http://www.orient.uw.edu.pl/~jsbien/>

Zakład Zastosowań Informatycznych

<http://www.orient.uw.edu.pl/~zzi/>

Instytut Orientalistyczny

<http://www.orient.uw.edu.pl>

Uniwersytet Warszawski

<http://www.uw.edu.pl>