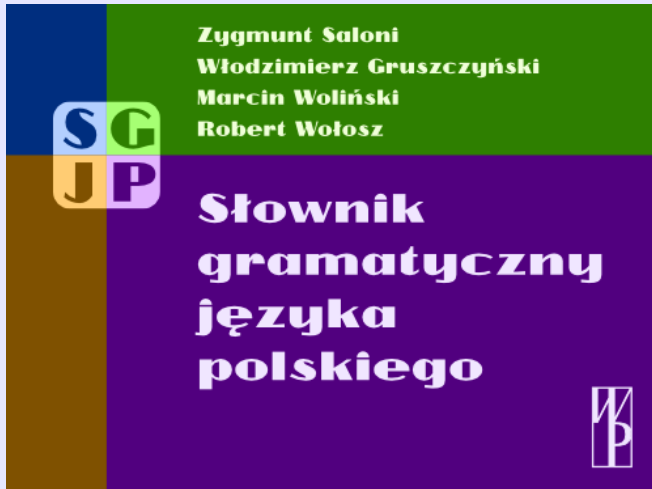




Marcin Woliński

**Model relacyjny
fleksji polskiej
w *Słowniku*
gramatycznym
*języka polskiego***



Warszawa, Wiedza Powszechna 2007



Publikacja przygotowana w latach 2003–2006
w ramach projektu
Słownik gramatyczny języka polskiego,
sponsorowanego przez Komitet Badań Naukowych
(nr rejestracyjny 2 H01D 007 24) i zlokalizowanego
w Uniwersytecie Warmińsko-Mazurskim w Olsztynie.



Zawartość słownika

Słownik gramatyczny języka polskiego

- podaje wyczerpującą informację o odmianie polskich leksemów (uwzględnia wszystkie wzorce odmiany obecne w języku?);
- uwzględnia podstawowe cechy składniowe:
 - dla rzeczowników — rodzaj,
 - dla czasowników — aspekt,
 - dla przymków — wymagany przypadek,
 - dla liczebników — typ związku z rzeczownikiem,
 - oraz wskazówki co do łączliwości leksemów nieodmiennych;
- nie zawiera definicji.



Zawartość słownika

	Hasła	Wzory
razem	244,341	1095
rzeczowniki	135,388	762
pospolite	68,099	
własne	8,754	
odstówniki	29,587	
odprzymiotnikowe	28,948	
przymiotniki	65,564	71
„prawdziwe”	31,273	
imiestowy	34,291	
liczebniki	98	45
czasowniki	29,531	215
nieodmienne	13,760	2
przysłówki	11,469	
przymyki	112	
inne	2,179	



Słownik gramatyczny języka polskiego

Słownik Lista hasel Odmiana Pomoc

Kaczyński    Wszystkie leksem  

kaczodziobosc	rz. f
kaczodzioby	przym.
kaczor	rz. m2
Kaczor	rz. m1
kaczorek	rz. m2
Kaczorowska	rz. f
Kaczorowski	rz. m1
kaczość	rz. f
kaczuchna	rz. f
kaczucza	rz. f
kaczusia	rz. f
kaczuszka	rz. f
kaczy	przym.
kaczyczka	rz. f
kaczyniec	rz. m3
Kaczyńska	rz. f
Kaczyński	rz. m1
Kada	rz. f
Kadafi	rz. m1
Kadar	rz. m1
kadaster	rz. m3
kadastralność	rz. f
kadastralny	przym.
kadastrować	cz. ndk
kadastrowanie	ods.
kadastrujący	im.cz.

244 366 hasel

Kaczyński

rzeczownik *n.* *wł. nazwisko* ◇ np.: Bogusław Kaczyński (publicysta i krytyk muzyczny)
m1 A708

	I. p.	I. m.	
M.	Kaczyński	Kaczyńscy	ndepr
		Kaczyńskie	depr
D.	Kaczyńskiego	Kaczyńskich	
C.	Kaczyńskiemu	Kaczyńskim	
B.	Kaczyńskiego	Kaczyńskich	
N.	Kaczyńskim	Kaczyńskimi	
Ms.	Kaczyńskim	Kaczyńskich	
W.	Kaczyński	Kaczyńscy	ndepr
		Kaczyńskie	depr

Odsyłacze

wariant żeński **Kaczyńska**



Praca nad SGJP

- Dane dla poszczególnych części mowy są przechowywane w osobnych bazach programu MS Access. Za każdą bazę jest odpowiedzialny jeden z Autorów.
- Wszystkie bazy mają (nieomal) tę samą strukturę.
- Przetwarzanie do postaci danych wykorzystywanej w interfejsie Słownika odbywa się pod Linuxem (mdb-tools, Perl i SQLite).



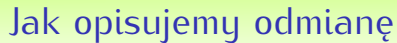
Jak opisujemy odmianę

	sg	pl
nom	dziecko	dzieci
gen	dziecka	dzieci
dat	dziecku	dzieciom
acc	dziecko	dzieci
inst	dzieckiem	dziećmi
loc	dziecku	dzieciach
voc	dziecko	dzieci



Jak opisujemy odmianę

	sg	pl
nom	dziecko	dzieci
gen	dziecka	dzieci
dat	dziecku	dzieciom
acc	dziecko	dzieci
inst	dzieckiem	dziećmi
loc	dziecku	dzieciach
voc	dziecko	dzieci

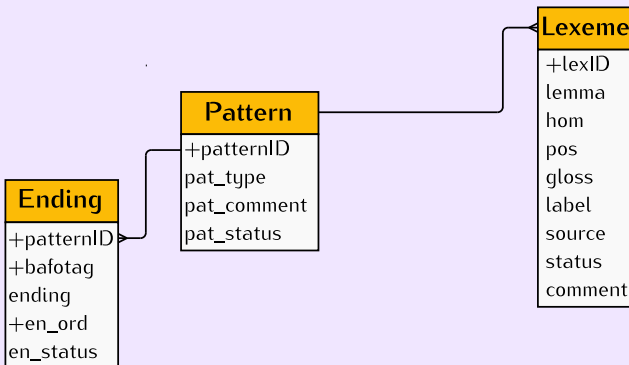


Wzór fleksyjny

dzie +		sg	pl
	nom	cko	ci
	gen	cka	ci
	dat	cku	ciom
	acc	cko	ci
	inst	ckiem	ćmi
	loc	cku	ciach
	voc	cko	ci

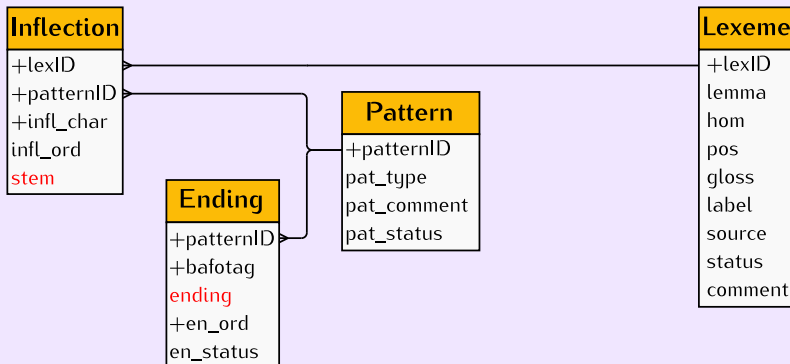


Schemat bazy danych (1. przybliżenie)





Schemat bazy danych (2. przybliżenie)





Odmiana typowego przymiotnika

	sg					pl	
	m1	m2	m3	n1,n2	f	m1,p1	pozostałe
nom	zły	zły	zły	złe	zła	źli	złe
gen	złego	złego	złego	złego	złej	złych	złych
dat	złemu	złemu	złemu	złemu	złej	złym	złym
acc	złego	złego	zły	złe	złą	złych	złe
inst	złym	złym	złym	złym	złą	złymi	złymi
loc	złym	złym	złym	złym	złej	złych	złych



Odmiana typowego przymiotnika

	sg					pl	
	m1	m2	m3	n1,n2	f	m1,p1	pozostałe
nom	zły	zły	zły	złe	zła	źli	złe
gen	złego	złego	złego	złego	złej	złych	złych
dat	złemu	złemu	złemu	złemu	złej	złym	złym
acc	złego	złego	zły	złe	złą	złych	złe
inst	złym	złym	złym	złym	złą	złymi	złymi
loc	złym	złym	złym	złym	złej	złych	złych

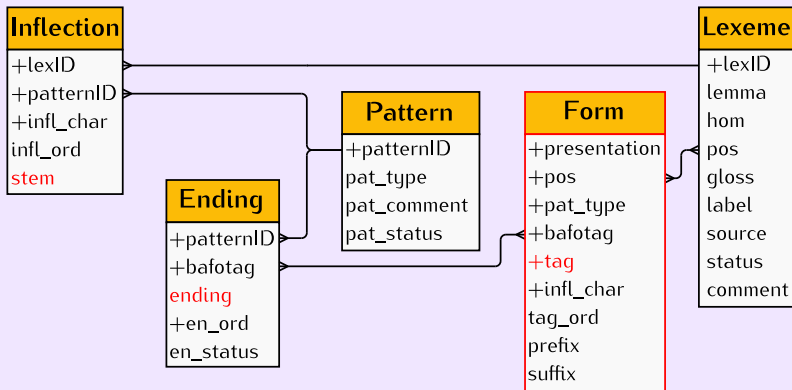


11 form bazowych przymiotnika

	sg					pl	
	m1	m2	m3	n1,n2	f	m1,p1	pozostałe
nom	1	1	1	5	6	9	5
gen	2	2	2	2	7	10	10
dat	3	3	3	3	7	4	4
acc	2	2	1	5	8	10	5
inst	4	4	4	4	8	11	11
loc	4	4	4	4	7	10	10



Schemat bazy danych (3. przybliżenie)





Trzy rzeczowniki, które odmieniają się „tak samo”

	mag m1		żuk m2		sąg m3	
	sg	pl	sg	pl	sg	pl
nom	mag	magowie magi	żuk	żuki	sąg	sągi
gen	maga	magów	żuka	żuków	sąga	sągów
dat	magowi	magom	żukowi	żukom	sągowi	sągom
acc	maga	magów	żuka	żuki	sąg	sągi
inst	magiem	magami	żukiem	żukami	sągiem	sągami
loc	magu	magach	żuku	żukach	sągu	sągach
voc	magu	magowie magi	żuku	żuki	sągu	sągi



Trzy rzeczowniki, które odmieniają się „tak samo”

	mag m1		żuk m2		sąg m3	
	sg	pl	sg	pl	sg	pl
nom	mag	magowie magi	żuk	żuki	sąg	sągi
gen	maga	magów	żuka	żuków	sąga	sągów
dat	magowi	magom	żukowi	żukom	sągowi	sągom
acc	maga	magów	żuka	żuki	sąg	sągi
inst	magiem	magami	żukiem	żukami	sągiem	sągami
loc	magu	magach	żuku	żukach	sągu	sągach
voc	magu	magowie magi	żuku	żuki	sągu	sągi



Połączony wzór dla męskiego typu odmiany

sg:nom	Ø
sg:gen	a
sg:dat	owi
sg:inst	iem
sg:loc	u
sg:voc	u
pl:nom:m1	owie
pl:nom:m2	i
pl:gen	ów
pl:dat	om
pl:inst	ami
pl:loc	ach
pl:voc	owie

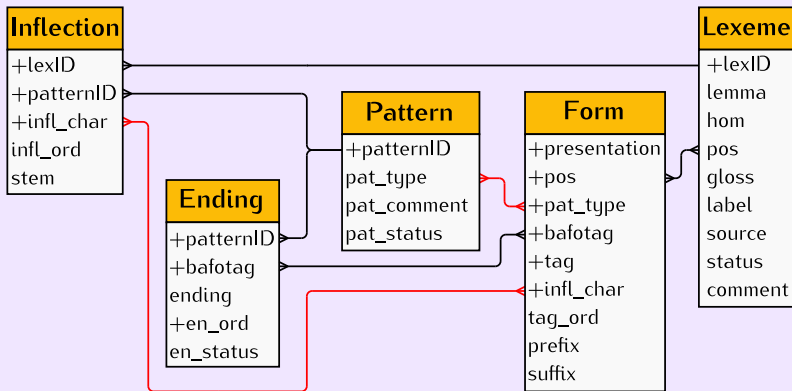


Typy rodzajowe odmiany rzeczowników

- Wskazane zależności między formami nie są uniwersalne dla wszystkich rzeczowników.
- Wprowadzamy typy rodzajowe wzorów rzeczownikowych: męski, żeński, nijaki, nieodmienny.
 - Encja Pattern ma atrybut `pat_type`.
- Generowane formy zależą także od rodzaju rzeczownika.
 - Encja Inflection ma atrybut `infl_char`.



Schemat bazy danych (4. przybliżenie)





Dlaczego infl_char jest atrybutem Inflection?

Ze względu na leksemy o rozchwianym rodzaju:

- CZŁOWIECZYSKO m1 lub n2
- CABERNET m2 lub m3 lub n2

Leksem taki opisujemy jedną instancją encji `Lexeme` skojarzoną z kilkoma instancjami `Inflection` o różnej wartości rodzaju.

Oczywiście dla typowych leksemów homonimicznych stosujemy osobne instancje encji `Lexeme` (np. trzy leksemy `BOKSER`).



Formy typowego czasownika

czytam, czytamy, czytasz, czytacie, czyta, czytają, czytałem, czytałam, czytałem, czytaliśmy, czytałyśmy, czytałeś, czytałaś, czytałoś, czytaliście, czytałyście, czytał, czytała, czytało, czytali, czytały, czytano, czytamy, czytaj, czytajcie, czytałbym, czytałabym, czytałobym, czytalibyśmy, czytałybyśmy, czytałbyś, czytałabyś, czytałobyś, czytalibyście, czytałybyście, czytałby, czytałaby, czytałoby, czytaliby, czytałyby, czytać, czytając, czytanie, czytania, czytaniu, czytaniem, czytań, czytaniom, czytaniami, czytaniach, czytany, czytanego, czytaniem, czytany, czytane, czytana, czytanej, czytana, czytani, czytanych, czytany, czytający, czytającego, czytającemu, czytającym, czytające, czytająca, czytającej, czytającą, czytających, czytającymi



Formy typowego czasownika

czytam, czytamy, czytasz, czytacie, czyta, czytają, czytałem, czytałam, czytałem, czytaliśmy, czytałyśmy, czytałeś, czytałaś, czytałoś, czytaliście, czytałyście, czytał, czytała, czytało, czytali, czytały, czytano, czytamy, czytaj, czytajcie, czytałbym, czytałabym, czytałobym, czytalibyśmy, czytałybyśmy, czytałbyś, czytałabyś, czytałobyś, czytalibyście, czytałybyście, czytałby, czytałaby, czytałoby, czytaliby, czytałyby, czytać, czytając, czytanie, czytania, czytaniu, czytaniem, czytań, czytaniom, czytaniem, czytaniach, czytany, czytanego, czytaniem, czytany, czytane, czytana, czytanej, czytana, czytani, czytanych, czytany, czytając, czytającego, czytającemu, czytającym, czytające, czytająca, czytającej, czytającą, czytających, czytającymi



12 form bazowych: czyta, czytam, czytają, czytaj, czytać, czytał, czytałem, czytało, czytali, czytano, czytanie, czytani



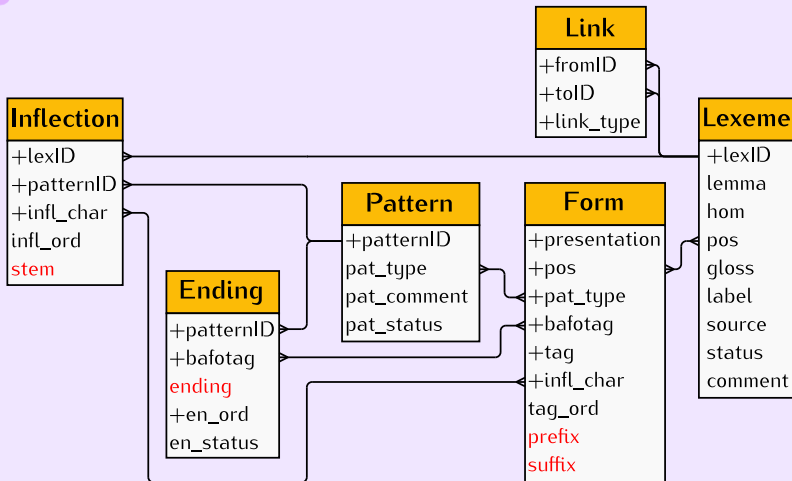
Części składowe czasownikowej formy fleksyjnej

prefiks
rdzeń
przyrostek tematowy
zakończenie
nry form bazowych

	d	ą		6	praet:sg:m:ter
	d	ą	em	7	praet:sg:m:pri
	d	ą	eś	7	praet:sg:m:sec
	d	ę	a	8	praet:sg:f:ter
	d	ę	o	8	praet:sg:n:ter
	d	ę	y	8	praet:pl:f:ter
	d	mą	cy	3	pact:sg:nom:m1:aff
nie	d	mą	cy	3	pact:sg:nom:m1:neg



Schemat bazy danych





Części składowe formy fleksyjnej

Form
prefix Inflection
stem

Ending
ending

Form
suffix Inflection
bafotag

Form

tag

naj	bia	tego	2	adj:sg:gen:m1
	biels	zego	2	adjsup:sg:gen:m1
	mag	a	sg:gen	sg:gen:m1
	mag	a	sg:gen	sg:acc:m1
	sąg	a	sg:gen	sg:gen:m3
	czyta	ł	8	praet:sg:n1:ter
	czyta	ł	8	praet:pl:f:ter
nie	czyta	n	10	ppas:sg:nom:m1:neg



Podsumowanie

- Przedstawiony model obejmuje wszystkie zjawiska fleksyjne uwzględnione w SGJP.
- Formy odmiany leksmu dowolnej klasy generuje się w ten sam sposób.
- Sposób prezentacji paradygmatów w programie można całkowicie zmienić manipulując tylko instancjami encji Form.