

Damian GRUND

Politechnika Śląska, Instytut Informatyki

KOMPUTEROWA IMPLEMENTACJA SŁOWNIKA SYNTAKTYCZNO-GENERATYWNEGO CZASOWNIKÓW POLSKICH

Streszczenie. W artykule opisano komputerową implementację słownika syntaktyczno-generatywnego czasowników polskich. Podano krótki wstęp do problematyki lingwistyki komputerowej, przedstawiono opis słownika w postaci tradycyjnej, opisano organizację bazy danych zawierającej informacje o czasownikach. Określono notację opisów syntaktycznych, podano przykłady zapisów haseł słownikowych w bazie danych oraz przedstawiono możliwości zastosowania słownika¹.

COMPUTER IMPLEMENTATION OF SYNTACTIC-GENERATIVE DICTIONARY OF POLISH VERBS

Summary. In this article computer implementation of dictionary of Polish verbs is described. A short introduction to problems of computer linguistics is presented as well as description of dictionary in traditional form, description of database organization containing information about verbs also has been given. Notation of syntactic description is shown together with examples of records regarding of dictionary entries in the database. Possibilities of dictionary application are also presented.

1. Wstęp

W dobie intensywnego rozwoju techniki zwiększa się zapotrzebowanie na informacje, co pociąga za sobą wzrost znaczenia nowoczesnych technik przesyłu i przetwarzania danych, w

¹ Opracowanie wykonano w ramach projektu badawczego 8 T11C 007 17 pt. „Translacja tekstów w języku polskim na język migowy”.

tym także inżynierii języka. Świadczy o tym duże zainteresowanie, jakim cieszą się ostatnio metody przetwarzania języka naturalnego.

Aby pewien język mógł zostać poddany przetwarzaniu komputerowemu, musi zostać opracowany zbiór cech opisujących ten język, czyli musi zostać opracowana jego gramatyka formalna [1].

W przypadku języków formalnych, takich jak na przykład języki programowania, problem ten praktycznie nie istnieje, gdyż języki te powstały w oparciu o opracowane wcześniej gramatyki formalne. Pozostaje zatem tylko problem implementacji komputerowej algorytmów zdefiniowanych przez daną gramatykę [2].

Inaczej wygląda sytuacja w przypadku języków naturalnych. Języki te powstały w sposób spontaniczny i podlegają ciągłemu rozwojowi, którego kierunku nie sposób określić. Pozostaje zatem problem opracowania gramatyki opisującej dany język i uwzględniającej jego twórczy charakter. Mimo że na razie nie udało się utworzyć gramatyki formalnej w pełni opisującej język naturalny, ciągle powstają gramatyki opisujące mniej lub bardziej dokładnie jego modele.

2. Opis słownika w postaci drukowanej

Słownik generatywny czasowników polskich jest narzędziem specjalizowanym, odzwierciedlającym strukturę składniową zdania. Obejmuje on cały zasób leksykalny języka oraz niezbędne informacje semantyczno-gramatyczne, umożliwiające tworzenie poprawnych zdań w języku polskim. Znaczenie zdania nie zależy tylko od znaczenia poszczególnych jednostek leksykalnych, ale także sposobu ich użycia, dlatego w słowniku [1] dla każdego czasownika określono zbiór schematów syntaktycznych, zawierających opisy struktur zdaniowych, w których może wystąpić. Poszczególne elementy struktury zdaniowej, związanej z danym czasownikiem, mogą zajmować tylko elementy o określonym znaczeniu, czyli należące do określonych klas leksykalno-semantycznych. Poniżej zostały podane przykłady jednostek słownikowych dla czasownika *zgotować* oraz czasownika *zaciekawić*, różniących się między innymi liczbą schematów syntaktycznych.

ZGOTOWAĆ

- $NP_N \rightarrow NP_{Acc} + NP_D$
 $NP_N \rightarrow [+Hum] [+Abstr]$
 $NP_{Acc} \rightarrow [+Abstr]$
 $NP_D \rightarrow [+Hum]$

Przykłady:

Zebrał widzowie zgotowali swemu ulubieńcowi gorące przyjęcie. – Syn zgotował mi piękną niespodziankę: ożenił się, nie pytając mnie o zdanie. – Nie wiesz jeszcze, jakie niespodzianki tydzień ci zgotuje. – Piękny las mu zgotował – powiedział Andrzej z wyrzutem. – Brak rozważania politycznego w przyszłości zgotuje narodowi serię kłesk.

Rys. 1. Sposób umieszczenia czasownika *zgotować* w słowniku [1]Fig. 1. The way of record of verb *zgotować* in the dictionary [1]

ZACIEKAWIAĆ - ZACIEKAWIĆ

1. $NP_N^1 \rightarrow NP_{Acc}^1 + (NP_P)$
 2. $NP_N^2 \rightarrow NP_{Acc}^2$
 $NP_N^1 \rightarrow [+Anim]$
 $NP_{Acc}^{1,2} \rightarrow [+Hum]$
 $NP_P \rightarrow [+Abstr]$
 $NP_N^2 \rightarrow \left[\begin{array}{l} -Abstr \\ -Anim \end{array} \right] [+Abstr]$

Przykłady:

1. Widać zaciekawia go nader ta dziewczyna.
 – Pies swym dziwnym zachowaniem zaciekawil nas wszystkich. – Pan mnie zaciekawia.
 2. Zaciekawilo mnie jego dziwne zachowanie. – Milicjanta zaciekawil porzucony na poboczu samochód.

Rys. 2. Sposób umieszczenia czasownika *zaciekawiać* w słowniku [1]Fig. 2. The way of record of verb *zaciekawiać* in the dictionary [1]

Słownik charakteryzując łączliwość poszczególnych czasowników, czyli możliwość jego wystąpienia w różnych kombinacjach wyrazowych, nie ogranicza się do podania cech semantycznych wyrazów, z którymi mogą one wchodzić w określone kombinacje (np. osobowy charakter rzeczownika, zajmującego pozycję podmiotu przy czasowniku „*bronić się*”, natomiast abstrakcyjny charakter rzeczownika, którego w pozycji tej wymaga zwrot czasownikowy „*ulec odchyleniu*”), ale także rozróżnia jego łączliwość fakultatywną od obligatoryjnej. W pierwszym przypadku chodzi o składniki, które muszą wystąpić przy danym czasowniku, aby zdanie było kompletne. W drugim chodzi o składniki, które mogą, ale nie muszą być użyte z danym czasownikiem. W tym celu został zdefiniowany ściśle określony, skończony, ale i zapewniający możliwość edycji, zbiór cech semantycznych rzeczowników, którego wielkość ma istotny wpływ na jakość i użyteczność słownika.

W słowniku został uwzględniony problem zwrotności czasowników oraz to, czy czasownik występuje w formie dokonanej lub niedokonanej. Zawarte są także informacje o

różnych znaczeniach tego samego czasownika, schematach zdaniowych odpowiadających poszczególnym znaczeniom, charakterystyki semantyczne składników nominalnych, umożliwiające określenie ograniczeń łączliwości, informacje o stronie biernej, przykłady oraz frazeologizmy.

Słownik w postaci drukowanej nie nadaje się bezpośrednio do wykorzystania w procesie komputerowej analizy tekstu, dlatego niezbędne było zaprojektowanie jego komputerowej implementacji.

3. Opis bazy danych

Podstawowym celem badań opisywanych w niniejszym artykule było zaprojektowanie struktury bazy danych umożliwiającej przeniesienie zawartości słownika w postaci drukowanej do postaci komputerowej oraz jego dalszy rozwój. W procesie dekompozycji dla realizacji bazy danych zostały wybrane cztery tabele: *t_main*, *t_schem*, *t_semo* i *t_semc*. Tabela *t_main* zawierała podstawowe informacje o czasowniku, czyli jego nazwę, znaczenie oraz przykład zastosowania, w tabeli *t_schem* zostały zamieszczone opisy schematów syntaktycznych czasowników, rekordy tabeli *t_semc* zawierały opis semantyczny poszczególnych elementów schematu oparty na zestawie 16 cech, natomiast w tabeli *t_semo* znajdowały się opisy semantyczne schematów. Dokładny opis tabel znajduje się w [2].

Wadą takiego rozwiązania było to, że dla opisów poszczególnych schematów tabela *t_main* zawierała zbędne, powtarzające się hasło czasownika. Pojawiły się także pewne ograniczenia w możliwości dalszego rozwoju bazy. Problem ten rozwiązano poprzez wprowadzenie dodatkowych tabel: *t_Link* łączącą czasownik z jego opisem syntaktycznym oraz tabelę *t_Desc* zawierającą znaczenie oraz przykład użycia danego czasownika. Obecnie baza opiera się na sześciu tabelach, jest łatwa w użyciu oraz w administracji, ma możliwość rozbudowy, a przede wszystkim ma możliwość przechowywania wszystkich niezbędnych informacji o czasownikach. Przyjęta struktura bazy jednoznacznie rozwiązuje problem alternatywy schematów syntaktycznych oraz koniunkcji i alternatywy cech i opisów semantycznych.

Zaimplementowany został także program umożliwiający administrowanie oraz wprowadzanie nowych haseł do bazy.

Poniżej podano szczegółowe opisy poszczególnych tabel.

3.1. Tabela *t_Main*

Tabela *t_Main* służy do przechowywania haseł słownikowych, czyli czasowników. Zawiera ona dwa pola: *Entry* i *Id*.

Pole *Entry* jest polem tekstowym indeksowanym, o unikatowych wartościach, co oznacza, że nazwa czasownika może wystąpić tylko jeden raz w całej tabeli. Czasowniki w tabeli umieszczane są w formie podstawowej, czyli w formie bezokolicznika. Aby umożliwić dalszy rozwój bazy oraz jednoznaczny dostęp do zawartości pól, czasowniki dokonane i niedokonane umieszczono osobno, mimo iż w formie drukowanej słownika [1] w większości przypadków posiadają takie same opisy.

Pole *Id* jest polem liczbowym, służy do połączenia tabeli *t_Main* z tabelą *t_Link*, relacją typu jeden-do-wielu. W procesie wprowadzania danych do bazy każdemu czasownikowi jest przyporządkowana unikalna w obrębie tabeli liczba całkowita. Dzięki takiemu rozwiązaniu następuje przyspieszenie wyszukiwania danych oraz możliwy jest dalszy rozwój bazy, polegający na dołączaniu nowych tabel.

3.2. Tabela *t_Link*

Tabela *t_Link* łączy czasownik z odpowiadającymi mu schematami syntaktycznymi i opisami semantycznymi. Pola tabeli nie zawierają bezpośrednich informacji dla użytkownika, niemniej jednak ma ona istotne znaczenie dla funkcjonowania całej bazy. W tabeli zostały zdefiniowane trzy pola: *Id*, *Nr_Schem* oraz *Nr_Desc*.

Pole *Id* jest polem liczbowym indeksowanym. Służy ono do połączenia tabeli *t_Main* z tabelą *t_Link* relacją typu jeden-do-wielu. Liczba rekordów o danej wartości pola *Id* jest równa liczbie schematów syntaktycznych, odpowiadających danemu czasownikowi.

Pole *Nr_Schem* jest polem liczbowym, indeksowanym, służącym do połączenia tabeli *t_Main* z tabelą *t_Schem* relacją typu wiele-do-jednego. Tak więc jest to pole łączące opisywane czasowniki z odpowiadającymi im schematami syntaktycznymi.

Pole *Nr_Desc*, podobnie jak pola *Id* oraz *Nr_Schem*, jest polem typu liczbowego, indeksowanym, zawiera wartości unikalne wewnątrz tabeli. Pole to pełni także funkcję klucza podstawowego. Służy ono do połączenia z tabelami *t_Desc*, *t_SchemAtt* i *t_SchemDesc* w oparciu o relacje typu odpowiednio jeden-do-jednego, jeden-do-wielu oraz jeden-do-wielu. Pole to jednoznacznie identyfikuje czasownik wraz z odpowiadającym mu opisem semantycznym, co umożliwi dalszy rozwój bazy, bez ingerencji w jej obecną wewnętrzną strukturę.

3.3. Tabela *t_Schem*

Tabela *t_Schem* służy do przechowywania schematów syntaktycznych. Rekordy tabeli zawierają dwa pola: *Nr_Schem* i *Schema*. Pole *Nr_Schem* jest polem liczbowym, zawierającym unikalny numer, służący do powiązania, poprzez tabelę *t_Link*, czasowników z

odpowiadającym im schematami syntaktycznymi. Pole *Schema* jest polem znakowym i zawiera zakodowany schemat syntaktyczny. Każdy schemat syntaktyczny składa się z ciągu ściśle zdefiniowanych symboli oraz znaków pomocniczych. Symbole składają się od jednej do sześciu liter, znaczenie poszczególnych symboli i znaków pomocniczych podano w tabeli 1. Schematy syntaktyczne wewnątrz tabeli nie powtarzają się, co znacznie ułatwia administrowanie bazą.

Tabela 1

Symbole i znaki pomocnicze

Symbol	Znaczenie
Adv	Przysłówek
Adj	Przymiotnik
Ip	fraza bezokolicznikowa (wyrażenie bezokolicznikowe)
K	partykuła pytajna „czy” lub zaimek pytajny (użyty w funkcji otwierania zdań podrzędnych pytajno-zależnych, np. „ <i>Nie wiem, kto to zrobił!</i> ”)
NA	rzeczownikowa nazwa czynności
NP	fraza nominalna (grupa rzeczownikowa)
NPabl	fraza nominalna ablatywna (tj. oznaczająca kierunek oddalania się, np. „ <i>z domu, ze wsi!</i> ”).
NPadl	fraza nominalna adlatywna (tj. oznaczająca kierunek przybliżania się, np. „ <i>do domu, do wsi!</i> ”).
NPakc	fraza nominalna akcesoryjna (tj. oznaczająca okoliczność towarzyszącą, np. „ <i>wśród burzy, w deszczu!</i> ”).
NPcaus	fraza nominalna przyczynowa (tj. oznaczająca przyczynę np. „ <i>z powodu kataru, z głodu, ze zmartwienia!</i> ”).
NPcond	fraza nominalna warunkowa (tj. oznaczająca warunek, np. „ <i>pod warunkiem dostarczenia towaru!</i> ”).
NPcons	fraza nominalna względu (np. „ <i>merytorycznie, formalnie!</i> ”).
NPdest	fraza nominalna celowa (np. „ <i>dla zabawy!</i> ”).
NPgrad	fraza nominalna stopnia i miary
NPloc	fraza nominalna lokatywna (miejscowa)
NPmod	fraza nominalna sposobowa
NPperl	fraza nominalna perlatywna (tj. oznaczająca miejsce, przez które odbywa się ruch, np. „ <i>przez las!</i> ”).
NPtd	fraza nominalna temporalno-duratywna (wyrażająca czas trwania)
NPtp	fraza nominalna temporalno-limitywna (wyrażająca granicę czasową)
OR	mowa wprost
Pass	strona bierna
S	Zdanie
Ts	korelat zaimkowy (<i>to, fakt</i> , np. „ <i>To, że on przyjdzie, jest pewne; Fakt, że tak się stało, dowodzi...</i> ”).
NP n,g,d,acc,i,l	małe litery użyte za symbolem frazy nominalnej i korelatu zaimkowego wskazują ich przypadek gramatyczny
pl	liczba mnoga

cd. tabeli 1

sg	liczba pojedyncza
„ - ”	pozycja czasownika hasłowego w schemacie zdaniowym
„ + ”	łączenie składników bez implikacji w szyku aktualnym zdania
„ ^ ”	konkatenacja (znak implikujący szyk składników)
„ () ”	fakultatywność składników lub grupy składników (tj. możliwość ich pominięcia w zdaniu)
„ {x,y,z} ”	wymienność składników w danej pozycji
znaki „ ”	wewnątrz tych znaków znajduje się fragment tekstu występujący zawsze w danej konstrukcji
„ ”	znak ten rozdziela składniki występujące wewnątrz nawiasów klamrowych (np. {NPn,Npacc,OR})

Dla przykładu, w oparciu o znaki i symbole zebrane w tabeli, został zbudowany schemat

$$Npn - Npacc + (Npd) \quad (1)$$

Oznacza on, że pierwszym elementem zwrotu jest fraza nominalna w mianowniku, drugim elementem jest dany czasownik, trzecim elementem jest fraza nominalna w bierniku, natomiast ostatnim - fraza nominalna w celowniku, której występowanie w poprawnym schemacie jest fakultatywne.

W słowniku znajdują się również schematy bardziej rozbudowane, na przykład schemat o postaci słownikowej:

$$Npn - Npacc + \left(\begin{array}{l} Npd \\ \text{"przed"} \wedge Npi \\ \text{"wobec"} \wedge Npg \end{array} \right) \quad (2)$$

Ze schematu (2) wynika, że dany czasownik jest poprzedzony frazą nominalną w mianowniku, a poprzedza frazę nominalną w bierniku. Aby zdanie było poprawne syntaktycznie, obydwie frazy muszą wystąpić na odpowiednich pozycjach, nawiasy zewnętrzne świadczą o tym, że wystąpienie frazy nominalnej w celowniku, frazy nominalnej w narzędniku, poprzedzonej przymikiem „przed” lub frazy nominalnej w dopełniaczu, poprzedzonej słowem „wobec”, jest możliwe, ale nie konieczne. Klamry wewnętrzne informują, że za frazą nominalną w bierniku może wystąpić dokładnie jeden z wymienionych elementów schematu. Z powyższego schematu wynika, że dany czasownik może wystąpić w następujących konstrukcjach syntaktycznych:

$NP_N - NP_{Acc}$

$NP_N - NP_{Acc} + NP_D$

$NP_N - NP_{Acc} + \text{przed} \wedge NP_i$

$NP_N - NP_{Acc} + \text{wobec} \wedge NP_G$

Schemat (2) został zamieszczony w bazie w następujący sposób:

$NPn - Npacc + \{ (Npd, \text{"przed"} \wedge Npi, \text{"wobec"} \wedge Npg) \}$

3.4. Tabela *t_Desc*

Tabela *t_Desc* służy do przechowywania informacji na temat znaczenia czasownika w określonym schemacie syntaktycznym, oraz przykładów zastosowania dla danego znaczenia. Dane zostały zawarte w trzech polach: *Nr_Desc*, *Desc* i *Example*.

Pole *Nr_Desc* jest polem liczbowym, indeksowanym, służy do powiązania tabeli *t_Desc* z tabelą *t_Link*. Spełnia ono także rolę klucza podstawowego tabeli. Pole *Desc* jest polem znakowym, zawiera krótki opis znaczenia danego czasownika. Pole *Example*, będące polem znakowym, zawiera przykład zastosowania danego czasownika w danym znaczeniu. Dane zawarte w tej tabeli mają jedynie znaczenie informacyjne, ze względu na brak formalnych zasad opisu nie nadają się do komputerowego przetwarzania, dlatego też informacje te nie znalazły się w osobnej tabeli, mimo iż istniała możliwość umieszczenia ich w poszerzonej tabeli *t_Link*.

3.5. Tabela *t_SchemAtt*

Tabela *t_SchemAtt* zawiera opisy semantyczne poszczególnych elementów schematów, oparte na zestawie szesnastu cech. Aby jednoznacznie powiązać opis semantyczny z elementem schematu, zdefiniowano następujące pola: *Nr_Desc*, *Nr_Symb* i 16 pól opisujących występowanie danej cechy. Informacje o istnieniu cech semantycznych, takich jak: *abstrakcyjność*, *konkretność*, *żywołność*, *nieżywołność*, *osobowość*, *nieosobowość*, *zbiorowość*, *żywiol*, *roślina*, *informacja*, *instytucja*, *narzędzie*, *płyn*, *maszyna*, *materiał* oraz *część* zostały umieszczone w polu typu logicznego. Wartość logiczna *Tak* informuje, że dany element schematu posiada taką cechę. Pole *Nr_Symb* typu liczba całkowita zawiera informacje, który element schematu (symbol) jest opisywany. Do powiązania z tabelą *t_Link* służy pole *Nr_Desc*.

3.6. Tabela *t_SchemschemDesc*

Tabela *t_SchemDesc* zawiera słowne opisy semantyczne poszczególnych elementów schematów. Dane te zostały umieszczone w polach: *Nr_Desc*, *Nr_Symb* i *SymbDesc*. Pola *Nr_Desc* oraz *Nr_Symb* pełnią te same funkcje co w tabeli *t_SchemAtt*. Dodatkowe pole *SymbDesc* typu znakowego zawiera opis słowny danego elementu schematu. W tabeli *t_SchemDesc* zamieszczono informacje semantyczne o elementach schematów, które nie dały się sklasyfikować za pomocą zdefiniowanych 16 cech semantycznych. Ze względu na brak sformalizowanych zasad opisu dane zawarte w tej tabeli nie nadają się do bezpośredniego wykorzystania w procesie komputerowej analizy mowy ludzkiej, lecz mogą stanowić podstawę dalszego rozwoju słownika.

4. Opis sposobu umieszczenia danych w bazie

Dane wprowadzono do bazy za pomocą interaktywnego programu działającego w środowisku Windows. Jak wspomniano, hasła czasowników umieszczono w tabeli *t_Main*. W trakcie umieszczania każdemu czasownikowi został przydzielony numer identyfikacyjny, zapisywany w polu *Id*. Następnie wypełniano tabelę *t_Link* w ten sposób, że dołączano do niej rekord, którego pole *Id* zawierało ten sam numer co pole *Id* w tabeli *t_Main*. W przypadku gdy danemu czasownikowi odpowiadał tylko jeden schemat, w polu *Nr_Schem* został zapisany jego adres, czyli numer w tablicy *t_Schem*. Jeśli czasownik posiadał kilka schematów, to tablica *t_Link* zawierała kilka rekordów z adresami odpowiednich schematów w tablicy *t_Schem*. W przypadku braku właściwego schematu w tablicy *t_Schem* do tej tablicy został dodany nowy rekord, zawierający odpowiedni schemat, a do pola *Nr_Schem* tabeli *t_Link* odpowiedni adres. Połączenie z tabelami *t_Desc*, *t_SchemAtt* i *t_SchemDesc* zrealizowano poprzez umieszczenie w polu *Nr_Desc* numeru identyfikującego opisy semantyczne. Jeżeli dany czasownik posiadał kilka opisów semantycznych nie wpływających na jego znaczenie, wystarczyło określić znaczenie dla jednego opisu, natomiast w przypadku gdy opisy semantyczne miały wpływ na znaczenie, to dla każdego opisu określono jego znaczenie oraz przykład zastosowania.

Poniżej zostały podane przykłady umieszczenia w bazie czasowników o różnych schematach syntaktycznych.

Przykład 1

W przykładzie tym opisano sposób umieszczenia czasownika *zachęcić*. Postać opisu tego czasownika w słowniku [1] przedstawiono na rys. 3. Czasownik ten ma dwa schematy syntaktyczne oraz powiązane z nimi opisy semantyczne.

ZACHĘCAĆ – ZACHĘCIĆ

1. NP_N^k — $NP_{Acc}^1 + \left\{ \begin{array}{l} \text{OR} \\ \text{do} \cap NP_D^1 \\ \text{zeby} \cap S \end{array} \right\} + (NP_I)$
 2. NP_N^k — $NP_{Acc}^1 + \left\{ \begin{array}{l} \text{do} \cap NP_D^1 \\ \text{zeby} \cap S \end{array} \right\}$
- NP_N^k — [+Hum]
 NP_{Acc}^1 — [+Hum]
 NP_D^1 — [+Abstr]
 NP_I — [+Abstr]
 NP_N^k — [+Abstr]

Przykłady:

1. Jedz, dziecko, dobre konfitury, dobre konfitury, sama smażyłam — zachęcała mnie babcia. — Anna bezskutecznie zachęcała syna do podjęcia studiów. — Alicja zachęcała gości do tego, żeby skorzystali z basenu. — Czarującym uśmiechem zachęcała nieśmiałego chłopca do wyznań.
2. Pogoda zachęcała do wyjazdu. — Mrok zachęcał do zwierzeń. — Powodzenie w publicystyce zachęciło go do publicznych odczytów. — Jej postępowanie zachęcało mnie do tego, żeby zaproponować wspólną wycieczkę.

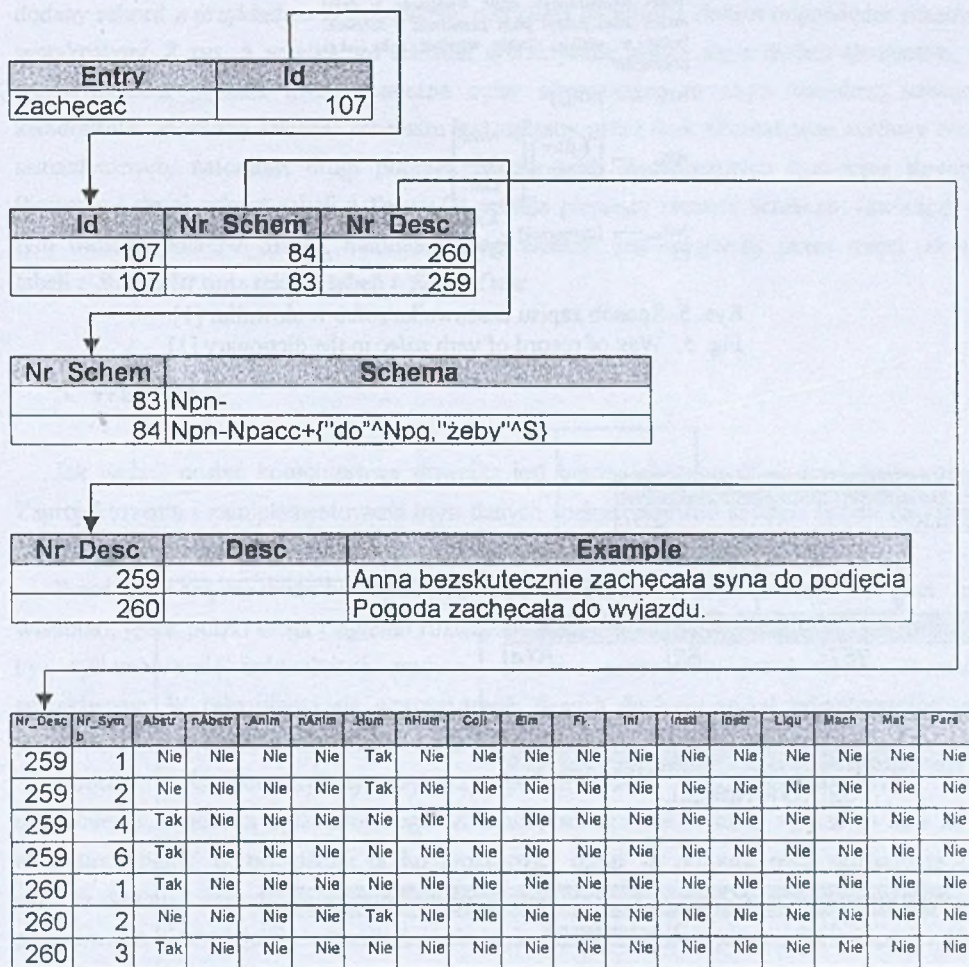
Rys. 3. Sposób zapisu czasownika *zachęcać* w słowniku [1]Fig. 3. Way of record of verb *zachęcać* in the dictionary [1]

Zgodnie z wcześniejszym opisem, dane opisujące czasownik zostały rozmieszczone w bazie w następujący sposób: (por. rys. 4)

1. Do tabeli *t_Main* dodano rekord zawierający w polu *Entry* nazwę czasownika, a w polu *Id* numer identyfikacyjny tego czasownika.
2. Czasownik posiada dwa niezależne schematy syntaktyczne, w związku z tym do tabeli *t_Link* dodano dwa rekordy o tym samym numerze *Id*, różniące się polami *Nr_Schem* oraz *Nr_Desc*. Jak widać, pola *Nr_Schem* zawierają liczby określające położenie odpowiednich schematów syntaktycznych w tabeli *t_Schem*.
3. Do tabeli *t_Desc* dodano dwa rekordy, zawierające przykład zastosowania czasownika w danych schematach. Pola *Desc* są puste, gdyż autor słownika [1] uznał, że znaczenie czasownika nie ma wpływu na rodzaj stosowanego schematu.
4. W tabeli *t_SchemAtt* umieszczono siedem rekordów, cztery opisujące pierwszy schemat oraz trzy drugi. Pole *Nr_Symb* określa, który element schematu jest obecnie opisywany, tak więc pierwszy rekord tabeli opisuje pierwszy symbol pierwszego schematu, drugi rekord opisuje drugi element pierwszego schematu, natomiast trzeci i czwarty opisują odpowiednio czwarty i szósty element pierwszego schematu. Odpowiada to zapisowi słownikowemu:

NP_N — [+HUM]
 NP_{Acc} — [+HUM]
 NP_G — [+ABSTR]
 NP_I — [+ABSTR]

W tabeli *t_SchemDesc* nie umieszczono rekordu, ponieważ nie zastosowano opisu słownego żadnego elementu schematu.



Rys. 4. Sposób rozmieszczenia danych o czasowniku *zachęcać* w bazie Fig. 4. Way of data distribution about verb *zachęcać* in the database

Przykład 2

Ten przykład zawiera opis sposobu umieszczenia czasownika *zalec*. Czasownik ten ma jeden schemat syntaktyczny, ale w przeciwieństwie do poprzedniego przykładu każdy element schematu posiada dwa różne opisy semantyczne. Opis czasownika w postaci słownikowej pokazano na rysunku 5.

ZALEC

'będąc umieszczonym gdzie, występując w dużej masie, ilości pokryć jakąś przestrzeń; o ludziach: będąc w wielkiej liczbie wypełnić sobą jakąś przestrzeń'

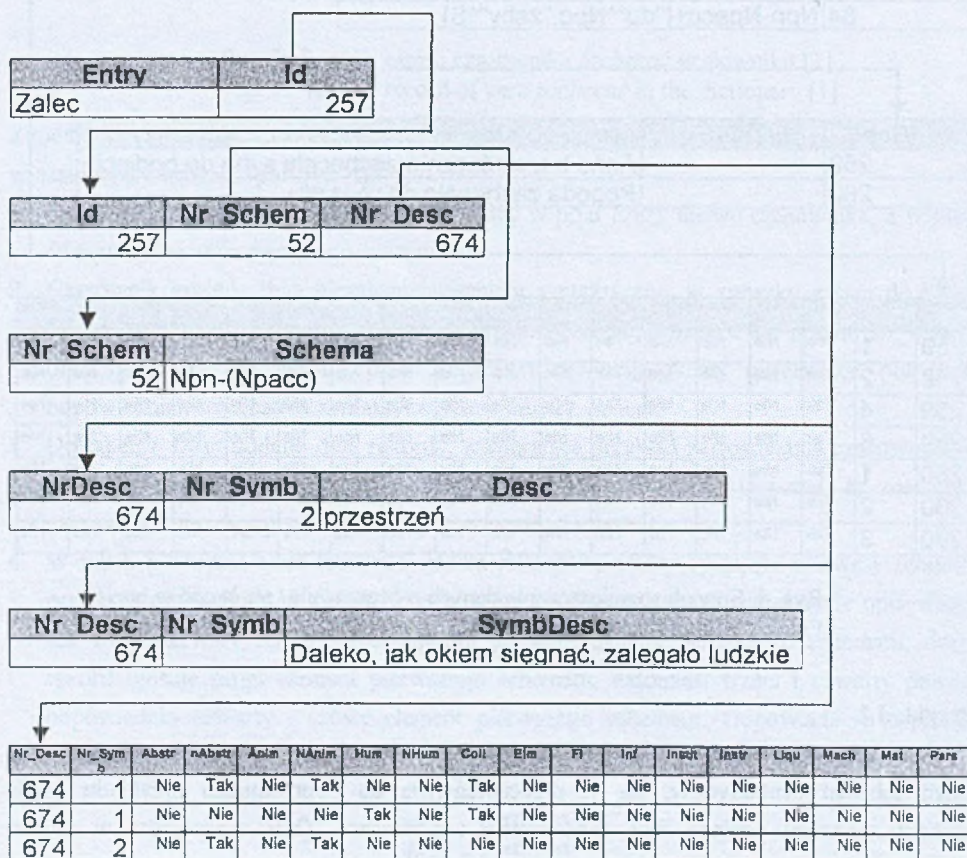
NP_u → (NP_{Acc})

NP_w → $\left[\begin{array}{l} +Hum \\ Coll \end{array} \right] \left[\begin{array}{l} -Abstr \\ -Anim \\ Coll \end{array} \right]$

NP_{Adv} → [przestrzeń] $\left[\begin{array}{l} -Abstr \\ -Anim \end{array} \right]$

Rys. 5. Sposób zapisu czasownika *zalec* w słowniku [1]

Fig. 5. Way of record of verb *zalec* in the dictionary [1]



Rys. 6. Sposób zapisu czasownika *zalec* w bazie

Fig. 6. Way of record of verb *zalec* in the database

Ponieważ czasownik ma jeden schemat syntaktyczny, w związku z tym w tabeli *t_Link* znajduje się tylko jeden rekord związany z tym czasownikiem. Do tabeli *t_Desc* został dodany rekord z przykładem zastosowania, natomiast do tabeli *t_Schem* odpowiedni schemat syntaktyczny. Z rys. 5 wynika, że schemat syntaktyczny składa się z dwóch elementów, a każdy element posiada dwa niezależne opisy semantyczne, o czym świadczą nawiasy kwadratowe. Pierwszy element schematu jest opisany przez dwa alternatywne zestawy cech semantycznych, natomiast drugi poprzez zestaw cech semantycznych oraz opis słowny. Pierwszy i drugi rekord tabeli *t_SchemAtt* opisuje pierwszy element schematu (świadczy o tym wartość pola *Nr_Symb*), natomiast drugi element jest opisywany przez trzeci rekord tabeli *t_SchemAtt* oraz rekord tabeli *t_SchemDesc*.

5. Wnioski

Jak widać, postać komputerowa słownika jest bardzo zbliżona do postaci drukowanej. Zaprojektowana i zaimplementowana baza danych spełnia zarówno kryteria narzucane przez ogólne zasady budowy baz danych, jak i założenia teorii gramatyki generatywnej.

Praca nad bazą nie może być ograniczona do komputerowej implementacji słownika. Jak wiadomo, język polski ulega ciągłemu rozwojowi, dlatego też, aby baza była użyteczna, musi być systematycznie uaktualniana zarówno o nowe czasowniki, jak i nowe schematy syntaktyczne. W celu ułatwienia wprowadzania danych do bazy został zaimplementowany program, umożliwiający dodawanie nowych rekordów przez osoby nie znające jej struktury.

Obecnie opis semantyczny wykorzystuje szesnaście cech semantycznych oraz uzupełnienie w postaci opisu słownego. Z analizy schematów wynika, że jest to opis zbyt mało precyzyjny. W przypadku braku możliwości opisu za pomocą cech semantycznych autorzy słownika zastosowali opis słowny, który ze względu na brak sformalizowanych reguł powstawania nie nadaje się bezpośrednio do wykorzystania w komputerze. W związku z tym zapewniono możliwość umieszczenia w bazie dodatkowych cech.

Zaprojektowana baza może być wykorzystana w różnych pracach związanych z przetwarzaniem języka, np. przy analizie lub generowaniu tekstów w języku polskim, w nauczaniu języka polskiego jako języka obcego, a także w różnego rodzaju aplikacjach, takich jak komputerowa korekta tekstów pisanych itp.

Aby umożliwić dostęp do zasobów słownika komputerowego „od zewnątrz”, czyli z programów ewentualnych użytkowników stworzonej bazy danych, opracowano bibliotekę klas dostępu w postaci kontrolki ActiveX, dzięki której to jest możliwe wykorzystanie bazy między innymi w aplikacjach internetowych. W dniu publikacji baza zawierała informacje o około 90% czasowników zapisanych w słowniku [1].

LITERATURA

1. Polański K.: Słownik syntaktyczno-generatywny czasowników polskich. Wyd. PAN, Warszawa 1980.
2. Grund D.: Opracowanie programu interaktywnego tworzenia bazy danych dla słownika semantycznego języka polskiego, praca magisterska, Politechnika Śląska, Gliwice 1998.
3. Vetulani Z., Abramowicz W., Vetulani G.: Język i technologia. Akademicka Oficyna Wydawnicza, Warszawa 1996.
4. Tokarski J.: Schematyczny indeks a tergo polskich form wyrazowych. Wyd. PWN, Warszawa 1993.
5. Świdziński M.: Gramatyka formalna języka polskiego. Wyd. UW, Warszawa 1992.
6. Bolc L., Cichy M., Różańska L.: Przetwarzanie języka naturalnego. Wyd. WNT, Warszawa 1982.

Recenzent: Dr hab. Zygmunt Vetulani Prof. UAM

Wpłynęło do Redakcji 12 września 2000 r.

Abstract

In this article computer implementation of Polish verbs in syntatic-generative dictionary is described. An introduction to computer linguistic is given with special attention to grammar describing natural languages is given. Basing on previously scanned dictionary entries description of dictionary in traditional form is presented. Organization of database of computer dictionary and description of separates tables it consists of is given. Symbols and auxillary marks are given as well as definition of notional regarding descriptions of syntatic schemes, the way of placing data in the database is presented together with examples of dictionary entries in classical form of dictionary and corresponding records in the table in the database. Possibilities of dictionary application and directions of its further development is also presented.