

WITOLD KIERAŚ, MARCIN WOLIŃSKI*

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA

Morfeusz 2 – analizator i generator fleksyjny dla języka polskiego¹

Słowa kluczowe: analiza i synteza morfologiczna, fleksja, przetwarzanie języka naturalnego.

1. Wprowadzenie

Analiza morfologiczna (fleksyjna), której celem jest przypisanie słowom tekstowym wszystkich możliwych interpretacji fleksyjnych, jest podstawowym zadaniem z zakresu automatycznego przetwarzania języka. Analizator nie dopasuje poprawnej interpretacji fleksyjnej do kontekstu składniowego czy semantycznego – tym zajmują się inne wyspecjalizowane programy² lub anotatorzy. Podstawowym celem analizatora jest zatem dostarczenie możliwie kompletnej informacji fleksyjnej anotatorom lub programom znakującym automatycznie zbiór tekstów w postaci elektronicznej.

Morfeusz oczywiście nie jest pierwszym programem przeznaczonym do analizy morfologicznej. Inne aplikacje tego typu powstawały co najmniej od połowy lat 90., czyli od ponad dwudziestu lat. Wśród nich warto wymienić m.in. warszawskie analizatory SAM autorstwa Krzysztofa Szafrana (1996) i AMOR Joanny Rabięgi-Wiśniewskiej i Michała Rudolfa (2002), poznańskie LEM (Vetulani i in. 1998) i UAM Text Tools (Obrębski, Stolarski 2005) czy też wciąż rozwijany na Węgrzech przez Roberta Wołosza PoMor (Wołosz 2005). Wszystkie te programy zostały szczegółowo omówione w przeglądowej pracy Elżbiety Hajnicz i Anny Kupść (2001). Do późniejszych aplikacji tego typu zaliczyć należy także analizator Morfologik (Miłkowski 2010), jak również obie wersje Morfeusza (pierwsza z nich została udostępniona w 2001 roku).

Od strony technicznej Morfeusz 2 to moduł programistyczny, który można wbudować w inne programy komputerowe lub wykorzystywać jako samodzielną aplikację. Jest on publicznie dostępny na bardzo liberalnej licencji BSD, umożliwiającej wykorzystanie programu w projektach zarówno naukowych, jak i komercyjnych. Program jest przygotowywany w postaci gotowej do użycia na popularnych platformach sprzętowych. Morfeusz jest zaawansowanym narzędziem informatycznym, wymagającym od użytkownika odpowiednich umiejętności

* wkieras@ipipan.waw.pl, wolinski@ipipan.waw.pl

1 Praca finansowana w ramach wkładu krajowego na rzecz udziału we wspólnym międzynarodowym przedsięwzięciu „CLARIN ERIC: Wspólne zasoby językowe i infrastruktura technologiczna”. Współautorstwo obejmuje wszystkie etapy pracy nad artykułem. Głównym twórcą koncepcji Morfeusza jest Marcin Woliński, głównym wykonawcą (programistą) Michał Lenart. Witold Kieraś jest autorem zestawu reguł segmentacyjnych analizatora.

2 Moduł ujednoznaczniający kontekstowo interpretacje fleksyjne nazywa się żargonowo tagerem. Z wykorzystaniem Morfeusza opracowano kilka tagerów języka polskiego, w szczególności TAKIPI zastosowany do znakowania Korpusu IPI PAN oraz tager Pantera zastosowany do znakowania Narodowego Korpusu Języka Polskiego.

i spełniającym ściśle określone wymagania w zakresie wąsko określonej dziedziny, niemniej jednak dla słabiej zorientowanych technicznie użytkowników przygotowano również interfejs graficzny, dzięki któremu można nie tylko korzystać z samego analizatora, ale też tworzyć jego własne zmodyfikowane wersje. Z działaniem Morfeusza 2 można także zapoznać się dzięki internetowej wersji demonstracyjnej pod adresem <http://sgjp.pl/morfeusz/demo/>.

2. Segmentacja

Morfeusz przetwarza tekst w postaci elektronicznej, będący ciągiem znaków³, i dokonuje segmentacji, czyli podziału ciągów znaków na takie podciągi (segmenty, słowa tekstowe), którym można przypisać jakąś interpretację fleksyjną, a następnie przypisuje im wszystkie możliwe interpretacje. Interpretacją fleksyjną nazywamy trójkę, w której skład wchodzi: 1) segment, 2) forma hasłowa leksemu (lemat) i 3) znacznik zawierający informację na temat klasyfikacji gramatycznej leksemu oraz wartości kategorii gramatycznych. Klasyfikacja gramatyczna w warstwie językoznawczej opiera się w dużej mierze na pracach Zygmunta Saloniego i Marka Świdzińskiego (Saloni 1974; Saloni, Świdziński 2001), dodatkowo jednak używa również pojęcia fleksemu, stworzonego przez Janusza Bienia (1991) i wykorzystywanego później m.in. przy tworzeniu systemu znaczników fleksyjnych w korpusie IPI PAN (Woliński 2003) oraz – w nieco zmienionej postaci – w Narodowym Korpusie Języka Polskiego (Przepiórkowski i in. (red.) 2012). Nie będziemy tu opisywać szczegółów technicznych i teoretycznych związanych z segmentacją ponad to, co jest niezbędne do zaprezentowania możliwości analizatora Morfeusz 2. Czytelnika zainteresowanego szczegółami odsyłamy do wyżej wymienionych publikacji.

Najprostszy sposób segmentowania wypowiedzenia polega na podzieleniu go na słowa tekstowe od spacji do spacji. Jest to jednak rozwiązanie niesatysfakcjonujące. Co prawda przyjmujemy założenie, że interpretowane segmenty nie przekraczają granicy słowa tekstowego, czyli maksymalny ich zakres wyznaczają otaczające je spacje, z pewnością jednak nie jest to podział odpowiednio drobny. Oczywistym i banalnym przykładem mogą być znaki interpunkcyjne, które – zgodnie z zasadami polskiej interpunkcji – w wielu wypadkach nie są oddzielone od słowa, po którym występują. Nieco poważniejszy problem stanowią formy czasu przeszłego i trybu przypuszczającego czasowników. Rozważmy przykładowe zdania (1) i (2), które zasadniczo wyrażają to samo i są równie poprawne – w pierwszym jednak człon *ście* wchodzi w skład niepodzielnej formy czasownikowej, w drugim zaś – jest zapisany łącznie ze spójnikiem podrzędnym. Niekiedy tego typu przesunięcie jest obowiązkowe i powoduje, że zdanie (4) jest gramatyczne, a zdanie (3) – nie.

- (1) Powiedziała, że to czytaliście.
- (2) Powiedziała, żeście to czytali.
- (3) *Powiedziała, żeby to czytaliście.
- (4) Powiedziała, żebyście to czytali.

3 W całym tekście terminu *znak* używamy w znaczeniu technicznym. Nie chodzi więc o znaki językowe, ale takie, które są przedmiotem np. standardu Unicode.

Z tego powodu pierwsza wersja Morfeusza traktowała wszystkie formy typu *czytaliście* jako dwa segmenty, nazywane odpowiednio pseudoimiesłowem przeszłym i aglutynantem. Jako osobne segmenty traktowane były również połączenia spójników podrzędnych i aglutynantów. Analogicznie opisywano także formy trybu przypuszczającego (z uwzględnieniem trzeciego segmentu *by*). Jest to opis językoznawczo spójny i uzasadniony, jednak w niektórych zastosowaniach nadmiernie skomplikowany, dlatego nowy Morfeusz umożliwił użytkownikowi wybranie interpretacji, w której formy czasu przeszłego i trybu przypuszczającego są interpretowane jako pojedyncze segmenty, jeśli tylko wszystkie człony tych złożonych form są zapisane łącznie.

Przesunięcie członu formy czasownikowej, będącego wykładnikiem osoby i liczby, nie ogranicza się jednak tylko do połączeń ze spójnikami podrzędnymi. Poniższe przykłady obrazują podobne tego typu połączenia z formami przymiotników (5), przysłówków (6), tradycyjnych zaimków rzeczownych (7) i rzeczowników (8). Takich połączeń analizator wcześniejszej wersji nie rozpoznawał, a obecnie jest to możliwe dzięki rozbudowanemu systemowi konstruowania reguł segmentacyjnych. Jak wszystko jednak, również analizowanie takich połączeń ma swoją cenę – zwiększa niejednoznaczność analiz, w skrajnych wypadkach jest źródłem systematycznej wieloznaczności, jak w połączeniach form męskich przymiotników w mianowniku (i bierniku – w wypadku podrodzaju nieżywołotnego) liczby pojedynczej z aglutynantem *-m*, będącym wykładnikiem pierwszej osoby liczby pojedynczej. Wtedy słowa tekstowe typu *ładnym*, *prostym* miałyby zawsze niejednoznaczne fleksyjne interpretacje, ponieważ mogą być to zarówno formy narzędnika liczby pojedynczej rodzaju męskiego i nijakiego, jak i formy przymiotnikowe z aglutynantem *-m*, przy czym ta druga interpretacja, choć możliwa, jest nieporównanie mniej prawdopodobna. Z tego powodu zatem możliwość analizowania takich połączeń jak w zdaniach (5)–(8) jest w Morfeuszu 2 fakultatywna i uruchamiana na życzenie użytkownika przez wybór odpowiedniej opcji w ustawieniach.

Co ciekawe, aglutynant może być również współdzielony przez dwie formy czasownikowe, jak w przykładzie (9), a to jest dodatkowym argumentem za uznaniem go za osobny segment.

(5) Możemy odpowiedzieć: a weźcie go sobie sami, skoro s a m i ś c i e go tu nam sprowadzili.

(W. Jagielski, *Modlitwa o deszcz*)

(6) Strasznie śmy się samym sobie podobali. (S. Aleksijewicz, *Czasy secondhand*⁴)

(7) Tam wszystkiego spróbowaliśmy, w s z y t k o ś m y poznali. (S. Aleksijewicz, *Cynkowi chłopcy*)

(8) No cóż, p r o f e s o r a m i ś m y, ale mamy i przezwiska komilitońskie jeszcze, burszowskie czasy pamiętające... (S. Lem, *Pamiętnik znaleziony w wannie*)

(9) W ciągu całej historii nie ż y l i ś m y, tylko s t a r a l i się przeżyć. (S. Aleksijewicz, *Czasy secondhand*)

3. Lematyzacja, czyli hasłowanie

Analizator segmentuje jednostki tekstowe w taki sposób, by mogły zostać zinterpretowane – każdemu z wyodrębnionych segmentów musi zatem być w stanie przypisać jakąś interpretację

4 Ten i kolejne cytaty z książek Swietłany Aleksijewicz w tłumaczeniu Jerzego Czecha.

fleksyjną. Interpretowanie polega na przypisaniu odpowiedniego lematu oraz znacznika fleksyjnego. Samo przypisywanie formy hasłowej nazywane jest lematyzacją lub hasłowaniem.

Lemat jest identyfikatorem leksemu, swego rodzaju konwencjonalną i arbitralnie dobraną etykietą, która jednoznacznie wyróżnia formy fleksyjne należące do paradygmatu odmiany tego samego leksemu. Tradycyjnie lemat jest wykładnikiem konwencjonalnie ustalonej formy leksemu, zwanej formą hasłową, np. dla czasowników jest to bezokolicznik. W wypadku jednak, gdy istnieje wiele leksemów o tej samej formie hasłowej, potrzebna jest dodatkowa informacja identyfikująca leksem. Taka sytuacja zachodzi np. dla dwóch leksemów o formie hasłowej *piec*: PIEC¹ (czasownik) i PIEC² (rzeczownik). Lematy tych dwu leksemów zostają w analizach Morfeusza rozszerzone dodatkowo o poprzedzoną dwukropkiem etykietę pozwalającą różnić te dwie jednostki: *v* dla czasownika i *s* dla rzeczownika. Jeśli homonimiczne jednostki reprezentują tę samą kategorię gramatyczną (czy ściślej: są opisane przez ten sam fleksem), przy etykietce pojawi się również indeks, np. rzeczowniki ADMIRAŁ¹ ‘stopień oficerski w marynarce wojennej’ i ADMIRAŁ² ‘gatunek motyla’, różniące się podrodzajem męskim, otrzymają odpowiednio formy hasłowe: *admiral:s1* i *admiral:s2*.

4. Przykład analizy

Rysunek 1 przedstawia przykładową analizę ciągu *Dziwny świat kota Filemona*. W kolumnie *Forma* widać kolejne wydzielone w wypowiedzeniu segmenty – w tym wypadku segmentacja jest prosta, jedynym segmentem wydzielonym nie na podstawie spacji między słowami jest końcowa kropka. W kolejnych wierszach znajdują się interpretacje fleksyjne dla poszczególnych słów tekstowych. Słowo *Dziwny* ma siedem interpretacji – we wszystkich wypadkach są to formy tego samego przymiotnika DZIWNY – które dzięki pewnej konwencji notacyjnej zostały skrócone do dwóch wierszy. Pierwszy wiersz mówi, że słowo może być formą biernika liczby pojedynczej rodzaju m₃ (tzw. męskiego nieżywoтного), drugi zaś – mianownika lub wołacza dowolnego z trzech rodzajów męskich. Kropka w zapisie typu *m1.m2.m3* jest alternatywą i pozwala skrócić trzy oddzielne interpretacje fleksyjne do jednego wiersza wynikowego.

Kolejny segment, czyli *świat*, ma dwie interpretacje rzeczownikowe: biernikową i mianownikową. Trzecie słowo (*kota*) może być formą jednego z trzech leksemów rzeczownikowych: mianownikiem żeńskiego rzeczownika KOTA, biernikiem lub dopełniaczem męskiego żywoтного rzeczownika KOT¹ ‘zwierzę’ albo biernikiem lub dopełniaczem męskiego osobowego rzeczownika KOT² ‘początkujący żołnierz, uczeń 1. klasy, student 1. roku’. W kolumnie *Kwalifikatory* widać, że interpretacje związane z rzeczownikiem KOT² są stylistycznie nacechowane, zostały oznaczone kwalifikatorami *pot.* (potoczne) i *środ.* (środowiskowe).

Słowo *Filemona* zostało rozpoznane jako biernik lub dopełniacz od imienia FILEMON. Ostatnim segmentem w analizie jest kropka, opisana po prostu jako znak interpunkcyjny. Ze względu na swoją wielofunkcyjność znaki interpunkcyjne w opisie fleksyjnym stosowanym przez Morfeusza na zasadzie konwencji są traktowane jako osobne segmenty (podobnie jak np. liczby zapisane cyframi rzymskimi lub arabskimi). Nie postulujemy oczywiście istnienia leksemów reprezentowanych przez znaki interpunkcyjne.

Analizator Morfologiczny Morfeusz

Analizator Generator

Słownik:
sgjp

Tekst:
Dziwny świat kota Filemona.

Analiza morfologiczna: Dopisz

O+	+O	Forma	Lemat	Tag	Nazwa	Kwalifikatory
0	1	Dziwny	dziwny	adj:sg:acc:m3:pos		
			dziwny	adj:sg:nom.voc:m1.m2.m3:pos		
1	2	świat	świat	subst:sg:acc:m3	nazwa pospolita	
			świat	subst:sg:nom:m3	nazwa pospolita	
2	3	kota	kota	subst:sg:nom:f	nazwa pospolita	
			kot:s1	subst:sg:acc:m2	nazwa pospolita	
			kot:s1	subst:sg:gen:m2	nazwa pospolita	
			kot:s2	subst:sg:acc:m1	nazwa pospolita [pot.; środ.]	
			kot:s2	subst:sg:gen:m1	nazwa pospolita [pot.; środ.]	
3	4	Filemona	Filemon	subst:sg:acc:m1	imię	
			Filemon	subst:sg:gen:m1	imię	
4	5	.	.	interp		

Zakończ Wyczyść Analizuj

Rysunek 1. Przykładowa analiza fleksyjna wypowiedzenia *Dziwny świat kota Filemona*

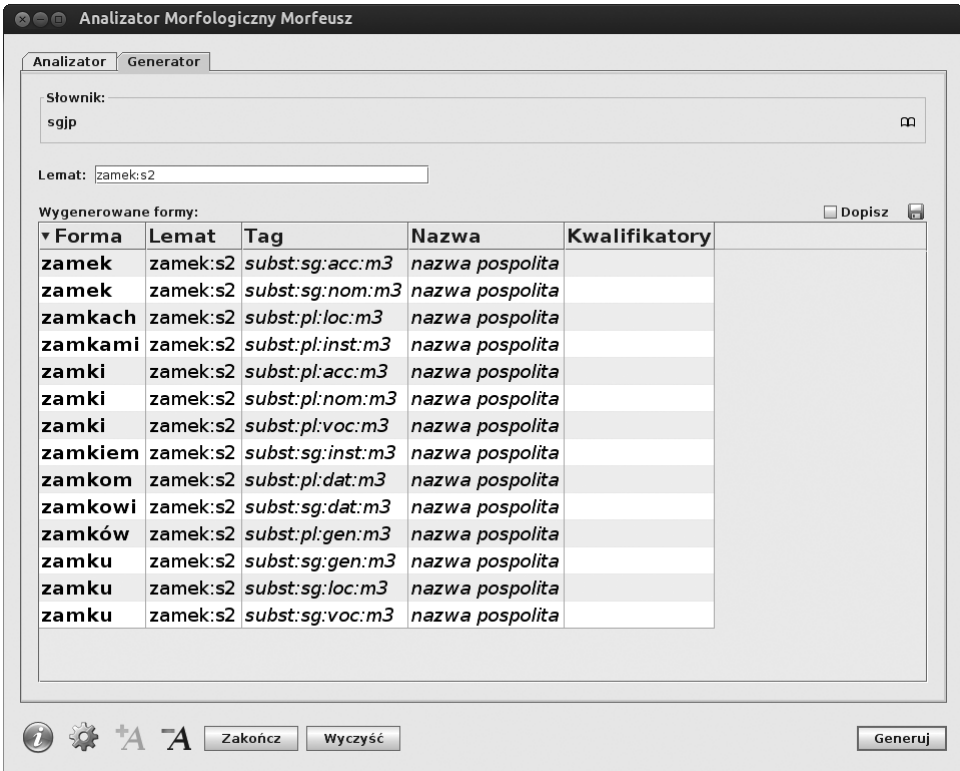
Jak widać, w kolumnie *Nazwa* wszystkie interpretacje rzeczownikowe mają przypisaną albo wartość *nazwa pospolita*, albo typ nazwy własnej, np. *imię*.

5. Generator

Zupełnie nową funkcją Morfeusza 2 jest możliwość generowania wszystkich form fleksyjnych leksemu. Do syntezy pełnych paradygmatów program używa tych samych danych słownikowych, które wykorzystuje do analizy.

W celu wygenerowania wszystkich form danego leksemu należy programowi podać jego formę hasłową. W wypadku form niejednoznacznych Morfeusz zwróci formy wszystkich leksemów o podanej formie hasłowej. Na przykład podanie programowi formy *zamek* spowoduje wygenerowanie dwu osobnych paradygmatów dla dwu leksemów: ZAMEK¹ ‘urządzenie’ i ZAMEK² ‘budowla’. Oba paradygmaty różnią się jedynie formą dopełniacza liczby pojedynczej, ale ponieważ w parze z różnicą w odmianie idzie również różnica znaczenia, stanowią one osobne hasła SGJP i – w konsekwencji – również w Morfeuszu. Ze względu jednak na homonimię form hasłowych będą one opatrzone identyfikatorami rozróżniającymi te leksemu. Jeśli użytkownik ma świadomość istnienia w słowniku więcej niż jednego leksemu o określonej formie hasłowej i chce w generatorze uzyskać formy tylko jednego z nich, może

programowi podać formę hasłową wraz z identyfikatorem, np. dla lematu *zamek:s2* generator zwróci wszystkie formy leksemu ZAMEK². Przykład syntezy form tego właśnie rzeczownika zaprezentowano na rysunku 2.



Rysunek 2. Przykładowa synteza wszystkich form jednego z dwu homonimicznych rzeczowników ZAMEK

6. Słowniki i segmentacja

Morfeusz dostępny jest z dwoma słownikami: aktualnymi danymi internetowej wersji *Słownika gramatycznego języka polskiego* (Saloni i in. 2015, por. artykuł w tym samym numerze) oraz słownikiem *Polimorf* (Woliński i in. 2012). Kolejne wydania Morfeusza są generowane automatycznie przez system Kuźnia (Szejko 2014) zarządzający pracą nad oboma słownikami. W przyszłości słowników może być więcej, ponieważ Morfeusz zaczyna być wykorzystywany również do analizy tekstów dawniejszych i może być rozszerzany o dane tworzone w innych projektach słownikowych (np. słownik fleksyjny polszczyzny XIX wieku po 1830 roku (Derwojedowa i in. 2014) czy słownik XVII i XVIII w. (Bronikowska i in. 2016)).

Możliwe jest również tworzenie różnych innych wyspecjalizowanych słowników fleksyjnych o konkretnym przeznaczeniu, które w specyficznych zastosowaniach albo całkowicie zastąpią standardowe słowniki, albo będą stanowić jedynie ich rozszerzenie. Przykładowo, można

sobie wyobrazić słownik fleksyjny obejmujący słownictwo typowe dla żargonu polskich internautów, stworzony z myślą o analizie korpusowej forów i czatów internetowych. Stanowiłby on rozszerzenie i uzupełnienie jednego z dwu standardowych słowników.

W wypadku Morfeusza słownik to po prostu duży plik tekstowy zawierający pięć kolumn. W każdym wierszu zapisana jest informacja o jednej formie fleksyjnej – w kolumnach znajdują się kolejno: forma tekstowa, lemat, znacznik fleksyjny oraz informacja o tym, czy jest to forma reprezentująca nazwę własną, i informacja o stylistycznych i zakresowych ograniczeniach użycia zapisana w postaci kwalifikatorów. Dwie ostatnie kolumny mogą być również puste.

Oprócz słownika do poprawnego działania Morfeusz potrzebuje jeszcze definicji tagsetu i reguł segmentacji. Tagset to zbiór wszystkich możliwych znaczników fleksyjnych, które analizator może przypisywać analizowanym segmentom. Z kolei w zbiorze reguł segmentacji – w pewnym uproszczeniu – definiuje się sposoby, w jakie ciąg znaków wejściowych może być podzielony na podciągi, by w efekcie powstała poprawna analiza. Wszystkie te trzy elementy, czyli słownik, tagset i reguły segmentacyjne, w nowej wersji Morfeusza można modyfikować i dostosowywać do potrzeb użytkownika. Najbardziej znacząca jest jednak możliwość tworzenia reguł segmentacji, otwierająca m.in. pole do wykorzystania pewnego rodzaju informacji słownikowej, która dotąd nie była wykorzystywana w analizatorze. *Słownik gramatyczny języka polskiego*, czyli podstawowe źródło danych lingwistycznych dla Morfeusza, notuje obok regularnych form deklinacyjnych przymiotników i liczebników również tzw. formy złożeniowe, czyli niesamodzielne jednostki występujące jedynie w określonych konstrukcjach słowotwórczych. Na przykład formy złożeniowe *popularno* i *pięcio*, będące formami odpowiednio przymiotnika POPULARNY i liczebnika PIĘĆ, występują w takich wyrazach jak POPULARNO-NAUKOWY, POPULARNOHISTORYCZNY, POPULARNOROZRYWKOWY; PIĘCIOCENTYMETROWY, PIĘCIOPIĘTROWY, PIĘCIORUBLÓWKA. Nie wszystkie z tych przykładowych wyrazów muszą się znajdować w słowniku analizatora, w szczególności rzadkie, choć poprawne, połączenia tego typu najpewniej w słowniku się nie znajdują, nie sposób bowiem wszystkich ich przewidzieć. Reguły segmentacyjne pozwalają jednak przekazać mechanizmowi analizatora informację, że jeśli dany ciąg znaków da się rozłożyć na dwa, z których pierwszy będzie formą złożeniową liczebnika lub przymiotnika, drugi zaś formą przymiotnika (czy w wypadku leksemu PIĘCIORUBLÓWKA – rzeczownika), to całość można zanalizować jako formę wyrazową o interpretacji fleksyjnej identycznej z interpretacją drugiego członu złożenia i formą hasłową, która składa się z formy hasłowej drugiego członu poprzedzonej odpowiednią formą złożeniową. Istnieje także możliwość takiego przekształcenia reguł segmentacyjnych, by powyższe przykładowe wyrazy były analizowane nie jako jeden segment, ale jako dwa oddzielne. Z możliwości tej korzysta się przy złożonych formach przymiotnikowych zapisywanych z dywizem typu *biało-czerwony*.

Tego samego mechanizmu używa się w Morfeuszu również w wypadku licznych przedrostków, takich jak *pseudo-*, *anty-*, *euro-*, *ekstra-*, *eks-*, które wchodzą w bardzo produktywne połączenia z rzeczownikami i przymiotnikami, tworząc wyrazy o łatwym do przewidzenia znaczeniu. Dzięki temu analizator jest w stanie poprawnie zinterpretować również rzadkie i nietypowe połączenia utworzone za pomocą przedrostków tego typu. W efekcie zatem Morfeusz

jest w stanie efektywnie analizować pewne określone produktywne konstrukcje słowotwórcze, choć oczywiście tylko nieliczne.

Podobnie jak w wypadku słowników, łatwo też można sobie wyobrazić różne zestawy reguł segmentacyjnych stosowane do różnych celów. Segmentacja we współczesnych tekstach literackich czy prasowych nie musi się pokrywać z segmentacją używaną do analizy tekstów dawniejszych, choćby ze względu na inne zasady pisowni łącznej i rozdzielnej. Na przykład dla słowników polszczyzny dawniejszej można sformułować regułę pozwalającą na analizowanie partykuły NIE zapisanej łącznie z formą czasownikową.

7. Podsumowanie

W artykule staraliśmy się w sposób przystępny przedstawić cele i zastosowania nowej wersji analizatora morfologicznego Morfeusz. Wszystkie opisane wersje analizatora dostępne są do pobrania na stronie <http://sgjp.pl/morfeusz/>. Nowe wersje są generowane raz w tygodniu, o ile w słowniku dokonano jakichś zmian. Dzięki temu dane Morfeusza są w zasadzie na bieżąco aktualizowane z internetową wersją SGJP.

Bibliografia

- Bień J.S. 1991: *Koncepcja słownikowej informacji morfologicznej i jej komputerowej weryfikacji*, Rozprawy Uniwersytetu Warszawskiego, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.
- Bień J.S., Saloni Z. 1982: *Pojęcie wyrazu morfologicznego i jego zastosowanie do opisu fleksji polskiej (wersja wstępna)*, „Prace Filologiczne” XXXI, s. 31–45.
- Bronikowska R., Gruszczyński W., Ogródniczuk M., Woliński M. 2016: *The use of electronic historical dictionary data in corpus design*, „Studies in Polish Linguistics”, 11(2), s. 47–56, doi:10.4467/23005920SPL.16.003.4818.
- Derwojedowa M., Kieraś W., Skowrońska D., Wołosz R. 2014: *Współczesne narzędzia leksykograficzne a analiza tekstów dawniejszych*, „Polonica” XXXIV, s. 21–27.
- Hajnicz E., Kupść A. 2001: *Przegląd analizatorów morfologicznych dla języka polskiego*, IPI PAN Research Report 937, Institute of Computer Science, Polish Academy of Sciences, Warsaw.
- Miłkowski M. 2010: *Developing an open-source, rule-based proofreading tool*, „Software: Practice and Experience”, 40 (7), s. 543–66.
- Obrębski T., Stolarski M. 2005: *UAM Text Tools – A text processing toolkit for Polish*, [w:] Z. Vetulani (red.), *Proceedings of 2nd Language & Technology Conference*, Wydawnictwo Poznańskie, Fundacja Uniwersytetu im. A. Mickiewicza, Poznań, s. 301–304.
- Przepiórkowski A., Bańko M., Górski R., Lewandowska-Tomaszczyk B. (red.) 2012: *Narodowy Korpus Języka Polskiego*, Wydawnictwo Naukowe PWN, Warszawa.
- Rabiega-Wiśniewska J., Rudolf M. 2002: *Amor – program automatycznej analizy fleksyjnej tekstu polskiego*, „Biuletyn Polskiego Towarzystwa Językoznawczego” LVIII, s. 175–186.
- Saloni Z. 1974: *Klasyfikacja gramatyczna leksemów polskich*, „Język Polski” LIV, s. 3–31, 93–101.
- Saloni Z., Gruszczyński W., Woliński M., Wołosz R., Skowrońska D. 2015: *Słownik gramatyczny języka polskiego*, wyd. 3 (online: <http://sgjp.pl>).
- Saloni Z., Świdziński M. 2001: *Składnia współczesnego języka polskiego*, Państwowe Wydawnictwo Naukowe, Warszawa.
- Szafran K. 1996: *Analizator morfologiczny SAM-95, opis użytkowy*, Rap. tech. TR 96-05 (226), Instytut Informatyki Uniwersytetu Warszawskiego, Warszawa.
- Szejko J. 2014: *Aplikacja webowa do opracowywania słowników fleksyjnych*, praca magisterska, Uniwersytet Warszawski, Warszawa.
- Vetulani Z., Martinek J., Obrębski T., Vetulani G. 1998: *Dictionary based methods and tools for language engineering*, Uniwersytet im. Adama Mickiewicza, Poznań.

- Woliński M. 2003: *System znaczników morfosyntaktycznych w korpusie IPI PAN*, „Polonica” XXII–XXIII, s. 39–55.
- Woliński M. 2006: *Morfeusz – a practical tool for the morphological analysis of Polish*, [w:] M.A. Kłopotek, S.T. Wierchoń, K. Trojanowski (red.), *Intelligent Information Processing and Web Mining*, Advances in Soft Computing, Springer, Berlin, s. 503–512.
- Woliński M. 2014: *Morfeusz reloaded*, [w:] N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (red.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, ELRA, Reykjavík, Iceland, s. 1106–1111.
- Woliński M., Miłkowski M., Ogrodniczuk M., Przepiórkowski A., Szałkiewicz Ł. 2012: *PoliMorf: a (not so) new open morphological dictionary for Polish*, [w:] N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (red.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012*, ELRA, Istanbul, Turkey, s. 860–864.
- Wołosz R. 2005: *Efektywna metoda analizy i syntezy morfologicznej w języku polskim*, Akademicka Oficyna Wydawnicza EXIT, Warszawa.

Summary

Morfeusz 2 – an inflectional analyser and generator for Polish

Keywords: morphological analysis and synthesis, inflection, natural language processing.

Morfeusz is a well known application, used in Polish computational linguistics for over a decade. In the paper we present a new version of the program, focusing on new features which have been introduced since the previous version.