

Katarzyna Głowińska

Taksonomia morfologiczna dla *Słownika frekwencyjnego*

Taksonomia morfologiczna została zaprojektowana jesienią 1999 r. (ostatnia modyfikacja z dnia 19 października 1999 r.) dla potrzeb analizy morfologicznej i anotacji wzbogaconego korpusu słownika frekwencyjnego. Inspiracją dla tej taksonomii był zestaw oznaczeń morfologicznych opracowanych dla języka czeskiego (Jan Hajič and Barbora Hladka. 1998. Tagging Inflective Languages: Prediction of Morphological Categories for a Rich, Structured Tagset. In *Proceedings of ACL/Coling'98*, Montreal, Canada, Aug. 5-9, pp. 483-490.)

Korpus *Słownika frekwencyjnego* został poddany działaniu analizatora morfologicznego SAM dr Krzysztofa Szafrana. Taksonomia ta miała więc na celu wykorzystanie informacji morfologicznej uzyskiwanej na wyjściu analizatora, a zarazem nadanie jej nowej formy. Istotne było zatem połączenie klasyfikacji na kategorie części mowy wykorzystywanych w *Słowniku frekwencyjnym* oraz w analizatorze SAM.

Oto przykładowy fragment oznakowanego korpusu:

B0001~Dziennik Bałtycki~05.01.1963~str. 5~kol. 1
Wychodząc[wychodzić,,V----W--N----P]
z[z,62,P-G-----P]
założenia[założenie,121,SSGN-----P] ,
że[że,,D-----P/C-----P]
różne[różny,212,APNOP-----P/APNRP-----P]
są[być,,VP---3TON----P]
właściwości[właściwość,112,SPNF-----P]
i[i,,C-----P]
różne[różny,212,APNOP-----P/APNRP-----P]
możliwości[możliwość,112,SPNF-----P]
zakładów[zakład,,SPGI-----P]
pracy[praca,121,SSGF-----P] ,
organizatorzy[organizator,,SPNP-----P/SPVP-----P]
konkursu[konkurs,,SSGI-----P]
przedstawili[przedstawić,,VP-0-3POD----P]

Każdy znacznik morfosyntaktyczny ma postać napisu złożonego z 14 znaków (tak samo dla wszystkich klas gramatycznych). Kategorie gramatyczne odpowiadające poszczególnym pozycjom zestawiono w poniższej tabeli. Jeżeli dana kategoria nie jest adekwatna dla danej klasy gramatycznej, na odpowiedniej pozycji stoi znak minusa.

Poz.	Znaczenie	Kod	Objaśnienie
1	klasa gramatyczna (część mowy)	V S A N Z D P C I T X	czasownik rzeczownik przymiotnik liczebnik zaimek przysłówek przyimek spójnik wykrzyknik partykuła kod nieznany
2	liczba	S P	pojedyncza mnoga
3	przypadek ¹	N G D A I L V	mianownik dopełniacz celownik biernik narzędnik miejsownik wołacz
4	rodzaj	M P A I F N O R T	męski męskoosobowy (l. poj.) męskozwierzęcy (l. poj.) męskorzeczowy (l. poj.) żeński nijaki męskoosobowy (l. mn.) niemęskoosobowy (l. mn.) plurale tantum
5	stopień	P C S	równy wyższy najwyższy
6	osoba lub oznaczenie formy bezosobowej czasownika	1 2 3 I B U W	pierwsza osoba druga osoba trzecia osoba bezokolicznik bezosobnik (forma na <i>-no</i> , <i>-to</i>) imiesłów przysłówkowy uprzedni imiesłów przysłówkowy współczesny
7	czas	T P F	teraźniejszy przeszły przyszły złożony
8	tryb	O P	oznajmujący przypuszczający

¹ Pozycja została także wykorzystana do przechowania informacji o wartości przypadkowej formy dla liczebników zbiorowych oraz kategorii przypadku form, z którymi łączą się przyimki.

Poz.	Znaczenie	Kod	Objaśnienie
		R	rozkazujący
9	aspekt	D	dokonany
		N	niedokonany
10	strona	C	czynna
		B	bierna
		Z	zwrotna
11	akcentowość	T	forma akcentowana
		N	forma nieakcentowana
12	poprzyimkowość	T	forma poprzyimkowa
		N	forma niepoprzyimkowa
13	oznaczenie dodatkowe form czasow- nikowych ²	I	bezokolicznik jako forma składowa czasu przyszłego (będzie <i>pisać</i>)
		S	forma na -ł jako składowa form czasu przyszłego (będzie <i>pisał</i>)
		P	forma na -ł jako czas przeszły z ruchomą końcówką (skoroś <i>zjadł</i>)
		W	forma trybu przypuszczającego z ruchomą partykułą (bym <i>napisał</i>)
		R	opisowa forma trybu rozkazującego trzeciej osoby (niech <i>pisze</i>)
		B	czasownik <i>być</i> jako składowa form czasów złożonych (<i>będzie pisał</i>)
		O	czasownik <i>być, bywać, zostać</i> jako składowe form strony biernej (<i>jest czytany, bywał sporządzany, zostanie zapisany</i>)
14	oznaczenie nazw własnych	P	nazwa pospolita
		W	nazwa własna
		S	skrótowiec

² Pole z oznaczeniami dodatkowych własności form czasownikowych zostało wprowadzone w celu reprezentacji obecnych na listach rangowych *Słownika frekwencyjnego* opisów funkcji czasownikowych form złożonych.