

Ludwik Wierzbowski

INFORMACJE

ZAWARTE W FORMACH FLEKSYJNYCH POLSKICH RZECZOWNIKÓW
ORAZ ICH ROZKŁAD NA TEMAT I KOŃCOWKĘ

Prace nad komputerowym przetwarzaniem danych językowych stworzyły potrzebę dwojakości opisu języka, a mianowicie rozróżnienia opisu syntezującego i analizującego. Opis syntezujący dokonywany jest ze stanowiska nadawcy. Na wejściu znajduje się informacja o tekście, na wyjściu - sam tekst. Innymi słowy: opisane jest przechodzenie od struktur głębszych do bardziej powierzchniowych, czyli po prostu - generowanie. Opis analityczny dokonywany jest ze stanowiska odbiorcy. Na wejściu znajduje się tekst, na wyjściu - zawarta w nim informacja. Tu opisane jest rozpoznawanie struktur głębokich na podstawie odpowiadających im powierzchniowych.

Tradycyjne opisy fleksji języka polskiego (jak zresztą w ogóle języków fleksyjnych) są typowo syntezujące, a jedynie zestawienia form zbieżnych (syncretic) w opisie deklinacji u Schenker¹ stanowią niejako próbę pogodzenia syntezy z analizą. Ten sam autor zresztą w pracy nad koniugacją² zajmuje stanowisko skrajnie generatywne.

Jako pierwszy w polskiej literaturze językoznawczej podejście analizujące, różne od syntezującego, zastosował w odniesieniu do fleksji J. Tokarski³, który funkcje końcówek fleksyjnych

¹ A. M. S c h e n k e r, *Polish Declension*, London-The Hague-Paris 1964.

² A. M. S c h e n k e r, *Polish Conjugation* [w:] "Word" 1954, t. X, s. 469-481.

³ J. T o k a r s k i, w serii artykułów pt. *Fleksja polska, jej opis w świetle możliwości mechanizacji w urządzeniu przekładowym*, zamieszczonych w

opisuje w sposób podobny do opisu np. formantów słowotwórczych, z myślą o przekładzie maszynowym. Nie wdając się w szczegółowy rozbiór ani w polemikę z tym, z czym się u Tokarskiego nie zgadza, rozpocznie autor własne rozważania od udowodnienia jego tezy, że analiza nie jest po prostu odwrotnością syntezy⁴.

Niech na jednym wejściu znajdzie się informacja fleksyjna 'G sg.' (*genetivus singularis*) oraz informacja słownikowa 'profesor'; na innym zaś wejściu - ta sama informacja fleksyjna obok słownikowej 'wykładowca'. Na odpowiednich wyjściach powinny się znaleźć formy fleksyjne *profesora* i *wykładowcy*. Kiedy jednak formy te znajdują się na wejściach to na odpowiednich wyjściach pojawiają się informacje słownikowe 'profesor' i 'wykładowca' (a więc zgodne z wejściami poprzedniej operacji), natomiast informacje fleksyjne będą: 'G sg. lub A sg.' i 'G sg. lub D sg. lub L sg. lub N. pl.' (a więc mniej dokładne niż na wejściach poprzedniej operacji i o niewielkim stopniu zgodności między sobą). Postawmy teraz na oba wejścia jednakową informację słownikową 'trupa' oraz dwie różne informacje fleksyjne: 'D sg.' i 'L sg.'. Stan obu wyjść będzie identyczny: *trupie*. Ta forma fleksyjna, postawiona na wejściu, da wyjściową informację: ('trupa' i 'D sg. lub L sg.') lub ('trup' i 'L sg. lub V sg.').

Z przykładów tych wynika, że 1) przetwarzanie informacji słownikowo-fleksyjnych na formy fleksyjne jest transformacją jednoznaczną jednokierunkowo, a nie wzajemnie jednoznaczna⁵; 2) przetwarzanie formy fleksyjnej na informacje słownikowe i fleksyjne jest transformacją niejednoznaczną, z czego wynika, że 3) informacje zawarte w formie fleksyjnej mają mniejszą (lub co najwyżej równą) wartość w porównaniu z informacją potrzebną do zbudowania tejże formy; 4) informacje zawarte w różnych formach fleksyjnych różnią się między sobą pod względem jakościowym i ilościowym.

w Por. Jęz. w latach 1961-1964. Późniejsza jego praca *Fleksja polska* (Warszawa 1973) w tej dziedzinie nic nowego nie wnosi.

⁴ "Rzecz nie polega na prostym odwróceniu wejść i wyjść automatów syntezy" (Por. Jęz. 1963, s. 3).

⁵ Terminy "transformacja jednoznaczna" (ang. single valued transformation), "transformacja jednoznaczna jednokierunkowo" (manyone transformation), "transformacja wzajemnie jednoznaczna (one-one transformation) pochodzą z pracy: W. R o s s A s h b y, *Wstęp do cybernetyki*, Warszawa 1963.

Ponieważ pierwszym etapem komputerowej analizy tekstu jest wydobywanie tych wszystkich informacji, jakie można uzyskać z pojedynczej, izolowanej formy wyrazowej z tekstu, trudno się oprzeć pokusie zastosowania miary informacji z wykorzystaniem matematycznej teorii komunikacji Shannona i Weavera⁶. Związkiem informacji fleksyjnych wydają się być łatwo wymiernymi. Dla rzeczowników polskich odmieniających się przez 7 przypadków w liczbie pojedynczej i przez 6 w mnogiej (bez *vocativu*) pełna i zarazem maksymalna informacja polegałaby na jednoznacznym wskazaniu przypadku i liczby (np. *pieśnią* - 'I sg.') i wynosiłaby w bitach $\log_2 13 - \log_2 1 = \log_2 13$ (jako że $\log_x 1 = 0$ bez względu na wartość podstawy logarytmu). Forma *pieśni* ('G sg. lub D sg. lub L sg. lub V sg. lub N pl. lub G pl. lub A pl.') zawierałaby informację fleksyjną wartości $\log_2 13 - \log_2 7$ bitów. Forma, która by dopuszczała wybór między wszystkimi 13 możliwościami, miałaby wartość informacyjną równą $\log_2 13 - \log_2 13 = 0$ bitów, taką samą jak forma nie wskazująca na żadną z 13 możliwości, która utrzymuje pełną entropię (stopień nieokreśloności), wynoszącą właśnie $\log_2 13$ bitów.

Jednakże taki rachunek byłby tylko wówczas realny, gdyby wszystkie z 13 form występowały z jednakową częstotliwością. Gdyby poszczególne częstotliwości były nam znane - a wiadomo, że są one nierówne - rachunek byłby znacznie bardziej zawiły. Dopóki jednak odpowiednimi danymi statystycznymi nie dysponujemy, możemy - i to w niektórych tylko wypadkach - stwierdzać jedynie równości i nierówności informacji fleksyjnych. Tak np. forma *człowiek* ('N sg.') zawiera więcej informacji fleksyjnych niż forma *zwierzę* ('N sg. lub A sg. lub V sg.'), za to forma *człowieka* ('G sg. lub A sg.') zawiera mniej informacji fleksyjnych niż forma *zwierzęcia* (wyłącznie 'G sg.'). Informacje fleksyjne form *sprawy* i *słowa* są równe ('G sg. lub N pl. lub A pl.'), a

⁶ Zasady teorii informacji zostały ogłoszone przez wspomnianych autorów w książce C. L. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Urbana 1949. Recenzja C. F. Hocketta w czasopiśmie "Language", nastawiona na lingwistyczne zastosowania teorii informacji, ukazała się również w przekładzie polskim J. Bronowicza w zbiorze *Z zagadnień językoznawstwa współczesnego*, Warszawa 1966. O zastosowaniu teorii informacji w językoznawstwie pisał także A. Bzdęga, *On the Linguistic Application of the Information Theory*, "Biuletyn Fonograficzny" 1965, t. VII, s. 13-24.

form *konie* i *wronie* nieporównywalne ('N pl. lub A pl.' wobec 'D sg. lub L sg.').

Formy fleksyjne rzeczowników polskich mogą zawierać 47 różnych informacji fleksyjnych, włączając do nich jedną formę zaimkową *mnie* ('G sg. lub D sg. lub A sg. lub L sg.') i 6 rzeczownikowych z końcówkami przymiotnikowymi: *wpisowe* ('N sg. lub A sg. lub N pl. lub A pl.'), *bliźni* ('N sg. lub V sg. lub N pl.'), *bliźnim* ('I sg. lub L sg. lub D pl.'), *zakopanem* ('I sg. lub L sg.'), *bliźnich* ('G pl. lub A pl. lub L pl.'), *królowych* ('G pl. lub L pl.'). Z pozostałych 40 informacji jedna występuje wyłącznie w formach nazw etnicznych *Cyganie* i *Hiszpanie* ('I sg. lub V sg. lub N pl.'), a dalsze 4 w formach pojedynczych rzeczowników: *ludu* ('G sg. lub V sg.'), *ręce* ('D sg. lub L sg. lub N pl. lub A pl.'), *dnie* ('L sg. lub N pl. lub A pl.') i *czasy* ('N pl. lub A pl. lub I pl.'). Pozostałe 35 informacji można uznać za regularne, z tym że różnią się one poważnie między sobą frekwencją w tekście oraz liczebnością leksemów, w których formach występują⁷.

Poważny błąd popełnia ten, kto posługuje się potoczną formułą, jakoby w formie fleksyjnej temat reprezentował treści znaczeniowe, końcówka zaś - gramatyczne (wyznacza stosunek do innych słów w wypowiedzeniu). Od błędy, z niejakimi zastrzeżeniami, można by ją częściowo zastosować do czasowników polskich, w żadnym natomiast wypadku do imion, a zwłaszcza do rzeczowników. Po pierwsze bowiem informacje o przypadku i liczbie nie wyczerpują całości informacji gramatycznych, gdyż niektóre z nich związane są z leksemem jako takim, nie zaś z określoną jego formą w tekście. Po wtóre kategorie przypadku i liczby (u przymiotników także rodzaju) wyznaczają, czy raczej określają, przynależność słowoformy do tej czy innej struktury składniowej w tekście oraz jej miejsce w tej strukturze takie czy inne, nie zawsze takie samo i nie zawsze w tym samym typie struktury, ale w zależności od właściwości gramatycznej samego danego leksemu i leksemów reprezentowanych przez inne słowoformy w tekście. Naiwnością więc byłoby twierdzić, iż czyni to bezpośrednio końcówka fleksyjna. Dlatego też tak powyżej, jak i w dalszym ciągu autor rozróżnia w

⁷ Szczegóły w obszernym - przygotowywanym przez autora do druku - opisie fleksji polskiej dla celów komputerowej syntezy i analizy tekstu.

formie fleksyjnej informacje fleksyjne i słownikowe (leksykalne), przy czym te ostatnie zawierają treści zarówno semantyczne, jak i gramatyczne, których rozróżnienia w tym miejscu nie podejmuje się.

Wbrew pozorom granica informacji słownikowej z fleksyjną wcale nie musi w każdym wypadku pokrywać się z granicą między tematem a końcówką, z czego gramatycy często nie zdają sobie sprawy (jak np. Schenker, gdy pisze: "Every paradigmatic form has a lexical meaning assigned to its stem and a grammatical meaning assigned to its ending"⁸), ale co w ich opisach wychodzi. Doroszewski⁹ wprost stwierdza, że należy brać pod uwagę całość formy, a nie poszczególne jej części. Że wymiana w obrębie tematu (niekoniecznie w języku alternacyjnym) może być wykładnikiem kategorii fleksyjnej, stwierdzono już dawno. Ale pomijając już nawet alternacje, istnienie podziału na deklinacje i koniugacje w gramatyce klasycznej świadczy, iż jej twórcy - przynajmniej intuicyjnie - wyczuwali, że np. różnica między łacińskim *-a* w *puella* i w *templa* ('N sg lub V sg.' wobec 'N pl. lub A pl.') kryje się w tematach tych rzeczowników.

Jak widzieliśmy w podanych wcześniej przykładach, zarówno fleksyjne, jak i słownikowe informacje mogą być (i nader często bywają) niepełne. Pełność informacji jest oczywiście pojęciem względnym. Ustaliśmy wyżej pełną informację fleksyjną polskiej formy rzeczownikowej, nie wnikając w to, czy pełność ta jest wystarczająca, bądź też w całości konieczna, dla zrozumienia przez odbiorcę tekstu, ani w to, czy i w jakim stopniu informacje niepełne mogą być uzupełniane przez kontekst. Nic nie stoi na przeszkodzie, by za pełną informację słownikową uznać umownie taką, która pozwala bezbłędnie i jednoznacznie rozpoznać leksem w takim sensie, w jakim np. pojmuje go Saloni¹⁰, nie wnikając zresztą zbytnio w jego subtelności semantyczne. Jeżeli któraś z tych informacji jest niepełna, możemy mówić o pozostawieniu przez formę pewnej ilości entropii fleksyjnej względnie słownikowej, bądź

⁸ Schenker, *Polish Declension*.

⁹ W. Doroszewski *Podstawy gramatyki polskiej*, Warszawa 1963.

¹⁰ Z. Saloni, *Klasyfikacja gramatyczna leksemów polskich*, Jęz. Pol. 1974, t. LIV, s. 3-13 i 93-101.

też jednej i drugiej. Otóż gdyby granica między tematem a końcówką zawsze się pokrywała z granicą między informacją słownikową a fleksyjną, to temat nie miałby żadnego wpływu na entropię i informację fleksyjną i *vice versa*. Przeczą temu następujące, bynajmniej nie unikalne zjawiska:

1. Informacje zawarte w temacie zmniejszają entropię fleksyjną, której utrzymaniu sprzyja homonimia końcówek. Tak np. końcówka *-u*, gdyby temat przy niej mógł być rozpoznany wyłącznie jako rzeczownikowy bez możliwości uzyskania jakiegokolwiek dalszej informacji słownikowej, dawałaby informację fleksyjną 'G sg. lub D sg. lub A sg. lub L sg. lub V sg. lub G pl.'. Tymczasem dla formy roku entropia fleksyjna ogranicza się do 'G sg. lub L sg. lub V sg.', dla formy wołu - do 'G sg. lub A sg.', dla formy ludu - do 'G sg. lub V sg.', dla formy morzu - do 'D sg. lub L sg.', dla formy widzu - do 'L sg. lub V sg.'. Formy łodu ('G sg.'), chłopu ('D sg.'), ciociu ('V sg.'), oczu ('G pl.') dają pełne informacje fleksyjne; tu entropia fleksyjna zostaje całkowicie skasowana.

W szczególnych wypadkach informacja fleksyjna zawarta jest w formie, która nie da się podzielić na temat i końcówkę, np. *tydzień* (podobnie jak łac. *iter*, *opus*, *Iuppiter* itp.). O końcówce zerowej nie może być tu mowy, musiałaby ona bowiem tworzyć opozycję z jedną co najmniej końcówką niezerową, gdy tymczasem formy takie, jak **tydzienia*, **iteris* itp. nie istnieją¹¹. W praktyce komputerowej takie bezkońcówkowe (fleksyjnie niepodzielne) formy wyrazów odmiennych mogą być dość liczne¹².

2. Informacje zawarte w końcówce zmniejszają entropię słownikową; a niekiedy nawet ją całkowicie likwidują. Tak np. formy *technikum*, *technikowi* i *technikę* dają - mimo homonimii tematów, która sprzyja utrzymywaniu entropii słownikowej - pełne informacje słownikowe: 'technikum', 'technik', i 'technika'.

3. Informacje zawarte w temacie i końcówce (niekiedy w samym tylko temacie) zmniejszają (lub nawet całkowicie likwidują) i jedną i drugą entropię; np. *kantory* to (('biuro handlowe' i 'N

¹¹ Por. L. Wierzbowski, *Nullmorphem oder Abwesenheit, eines Morphems in der komputationellen Grammatik einer natürlichen Sprache*, "Lingua Posnaniensis" 1977, t. XX, s. 121-126.

¹² Por. I. A. Mielczuk, *Morfologiczeskij analiz pri maszinnojm pieriewodie*, "Problemy Kibernetiki" 1961, t. VI, s. 209-276.

pl. lub A pl.'), podczas gdy *kantorzy* to ('śpiewak kościelny' i 'N pl.'). Inny przykład: *technice* to ('technika' i 'D sg. lub L sg.)'), natomiast *technicy* to ('technik' i 'N pl.'). W obu przykładach do zwiększenia obu informacji przyczyniły się alternacje w tematach, w drugim przykładzie także dobór końcówek.

4. Równoczesna homonimia tematu i końcówki sprzyja utrzymywaniu się obu entropii. Przykładem niech będzie forma *technika*, która daje informację ('technika' i 'N sg.') lub ('technik' i 'G sg. lub A sg.') lub ('technikum' i 'N pl. lub A pl.').

5. Entropia słownikowa i fleksyjna może ulec w wyjątkowych wypadkach przypadkowemu zwiększeniu bez homonimii tematu i końcówki, np. w formie *posłowie*, która daje informację ('posłowie' i 'N sg. lub A sg.') lub ('poseł' i 'N pl.').

6. Jeszcze bardziej zwiększa entropię homonimia tematów i końcówek między różnymi częściami mowy, zarówno homonimia typu 4 (np. *drogi*), jak i typu 5 (np. *pił-a* rzeczownik, ale *pił-a* czasownik).

Wszystkie te zjawiska stawiają pewne wymagania opisowi fleksyjnemu dla celów komputerowej analizy tekstu, jak np. konieczność zestawienia homonimii końcówek, odmiennej od tradycyjnej klasyfikacji fleksyjnej leksemów i in.¹³

Ciekawostką może być klasyfikacja końcówek rzeczownikowych ze względu na zawartość informacji fleksyjnej oraz na ich przydatność do zmniejszania entropii słownikowej. Można wśród nich wyróżnić 4 klasy:

1. Końcówki *-om*, *-ami* (z wariantem *-mi*), *-ach* dają z jednej strony pełną informację fleksyjną, z drugiej strony są absolutnie nieprzydatne dla zwiększania informacji słownikowej w obrębie samych rzeczowników. Dlatego też np. formy *technikom*, *technikami* i *technikach* pozostawiają wybór między wszystkimi trzema leksemami o tematach homonimicznych ('technika' lub 'technik' lub 'technikum'). W wyjątkowych wypadkach końcówki te mogą służyć odróżnieniu rzeczownika od przymiotnika, np. *łabędziom*, *łabędziami*, *łabędziach* wobec *łabędzim*, *łabędzimi*, *łabędzich*.

2. Końcówki *-owi*, *-e*, *-owie* (a także przymiotnikowa *-emu*) dają pełną informację fleksyjną i są zdolne do zmniejszenia entropii słownikowej. Porównaj np. przytoczone przykłady *technikowi* i

¹³ Por. przyp. 7.

technikę oraz różnicę między *zbiegowie* (np. polityczni) a *zbiegi* (np. ckoliczności).

3. Końcówka *-um* (a także przymiotnikowa *-ej*, np. *królowej*) dają informację małą (*-um* tylko o liczbie, nie o przypadku), ale za to niezależną od tematu i ze swej strony mogącą zmniejszać entropię słownikową. Porównaj przytoczony wyżej przykład *technikum*.

4. Wszystkie pozostałe - najliczniejsze - końcówki są w mniejszym lub większym stopniu homonimiczne i dają większe czy też mniejsze informacje, lecz zawsze zależne od tematu. Same też są predestynowane do uzupełniania informacji słownikowych (przynajmniej teoretycznie, gdyż dla niektórych może brakować przykładu). Najbardziej homonimiczne są: *-y* (16 różnych informacji), *-i* (14), *-u* (11), *-e* nie zmieniające tematu (9), *-e* z wymianą w temacie i *-a* (po 8)¹⁴.

Wiele z powyższych stwierdzeń dotyczących rzeczowników da się przenieść z grubsza na przymiotniki i inne wyrazy odmienne przez przypadki, liczby, rodzaje (różnice w znacznej mierze natury ilościowej). Natomiast trudno by je odnieść do czasowników, których analiza dużo bliższa jest odwrotności syntezy.

Uniwersytet Gdański
Gdańsk

Ludwik Wierzbowski

THE IMPLICATIONS OF INFLEXIONAL FORMS IN POLISH NOUNS; THEIR DIVISION INTO ROOTS AND ENDINGS

The author proves the necessity of two different language descriptions in computer synthesis and analysis of texts. Later he discusses the information carried by inflexional forms of Polish nouns from the point of view of its range and quality.

He distinguishes between lexical and inflexional information, its boundary being inconsistent with that of the root and the ending, since the information in the two is complementary as has been shown in the examples.

The last part of the paper reviews the characteristics of endings of Polish noun declensions.

¹⁴ *Ibid.*