

Konstrukcje werbalne z AŻ w gramatyce Świdzińskiego*

Janusz S. Bien[†]

23 lutego 2000

24 marca 2003

1 Wstęp

Przedstawione w tytule zagadnienie jest bardzo wąskie. W niniejszym artykule stanowi ono przede wszystkim pretekst do przedstawienia wyników analizatora syntaktycznego AS, opracowanego przez Marcina Wolińskiego w ramach kierowanego przeze mnie projektu KBN pt. *Zestaw testów do weryfikacji i oceny analizatorów języka polskiego*, zrealizowanego w Instytucie Informatyki Uniwersytetu Warszawskiego w latach 1997–1999. Jest to również okazja do sformułowania kilku uwag metodologicznych na temat tworzenia gramatyk formalnych dla języka polskiego.

W artykule stosowane są następujące konwencje notacyjne. Jednostki leksykalne zapisywane są zawsze kapitalikami, np. AŻ. Symbole gramatyki formalnej i elementy wyników analizy syntaktycznej zapisywane są czcionką przypominającą wydruk komputerowy, np. AŻ1. Wyrazy, frazy i zdania cytowane w tekście zapisywane są pismem pochylonym i ujmowane w nawiasy kątowe (co jest wygodne, gdy cytat zawiera również znaki interpunkcyjne), np. $\langle\langle$ Czeka, aż przyjdzie. $\rangle\rangle$.

2 Gramatyka Świdzińskiego

Przez gramatykę Świdzińskiego rozumiemy powierzchnię gramatyki języka polskiego sformułowaną w książce [12]. Była ona inspirowana przez wcześniejsze prace Stanisława Szpakowicza [10] oraz Szpakowicza i Świdzińskiego [11].

Gramatyka Szpakowicza, pisana od razu z myślą o jej implementacji w postaci analizatora syntaktycznego, miała dobrze zdefiniowaną postać tzw. gramatyki metamorficznej [5] (por. także [6]). Gramatyka Świdzińskiego w założeniu autora nie jest przeznaczona do bezpośredniego wykorzystania w zastosowaniach komputerowych, pod względem formalnym stanowi ona pewną modyfikację gramatyk metamorficznych. Ze względu na obszerność i złożoność tej gramatyki

*Niniejszy artykuł stanowi rozszerzoną wersję tekstu złożonego do tomu dedykowanego Prof. Marii Szupryczyńskiej. Niniejsza wersja elektroniczna jest dostępna na zasadach *GNU Verbatim Copying and Distribution*; numeracja stron różni się od wersji drukowanej.

[†]jsbien@mail.uw.edu.pl [adres z czasu publikacji, obecnie nieaktualny]

nie wszystkie konsekwencje tej modyfikacji są oczywiste, dosłowna interpretacja gramatyki prowadzi niekiedy do rezultatów niezgodnych z intencjami autora. Pomimo tej i innych odkrytych wad gramatyka ta stanowi osiągnięcie imponujące wewnętrzną konsekwencją i szczegółowością opisu.

3 Analizator syntaktyczny AS

Celowość komputerowej implementacji gramatyki Świdzińskiego była dla mnie oczywista od momentu jej powstania, jednak warunki do realizacji tego zadania powstały dopiero po pojawieniu się analizatora morfologicznego SAM opracowanego przez Krzysztofa Szafrana [8] (por. także np. [9]) na podstawie schematycznego indeksu polskich form wyrazowych [13]. Pierwsza próba skonstruowania analizatora syntaktycznego dla gramatyki Świdzińskiego została podjęta w ramach grantu KBN pt. *Analizator morfologiczno-syntaktyczny dla obszernego podzbioru języka polskiego*, zrealizowanego w latach 1994–1996 w Instytucie Informatyki UW pod moim kierunkiem. Choć początkowo działanie analizatora AMOS — przedstawione na LIV Zjeździe PTJ w Toruniu [2] (por. także [1]) — wydawało się zadowalające, okazało się później, że dla pełnej gramatyki jego szybkość działania jest zbyt niska, aby mógł on być praktycznie użyteczny.

Do problemu powrócono w grantie KBN wspomnianym we wstępie, którego sprawozdanie końcowe [3] jest dostępne w Internecie. Zrealizowany wówczas analizator AS potrzebował 1 sekundy — na komputerze PC średniej mocy — na wyprodukowanie jednej analizy dla typowego zdania [4], jednak dla wielu interesujących przykładów jego efektywność nie była wystarczająca. Jego autor, Marcin Woliński, kontynuował pracę nad analizatorem i obecnie¹ dysponuje on programem (nazwanym *Świgrą*) znacznie szybszym i wygodniejszym w użyciu. Zawarta w niniejszym artykule dyskusja odnosi się w jednakowym stopniu do analizatorów AS i Świgrą.

4 Konstrukcje werbalne właściwe

Skoncentrujemy się tutaj na tylko jednym typie konstrukcji werbalnych opisywanych przez gramatykę Świdzińskiego, a mianowicie na konstrukcjach werbalnych właściwych. Odpowiednie reguły gramatyki z sekcji 5.10.5 Aneksu w książce [12] (s. 380), po naniesieniu oczywistych poprawek, mają następującą postać.

```
KWER1 (wf, a, c, t, rl, o, wa, wb, wc, k, i, z)
= FORMACZAS (wf, a, c, t, rl, o, wa, wb, wc, k, i, z) (WE30)
  $ RÓWNE (z, NP.NP'.NP")
= FORMACZAS (wf, a, c, t, rl, o, wa, wb, wc, k, i, z) (WE31)
  $ RÓWNE (z, NPt.NP't.NP"t.P.P'.P".PZ)
  $ RÓŻNE (t, ROZ)
= FORMACZAS (wf, a, c, t, rl, o, wa, wb, wc, k, i, z) (WE32)
  $ RÓWNE (z, AŻ1".AŻ2.BO.BO'.BO".BOWIEM.CHOCIAŻ.CO.CZY.
    DOPÓKI.GDY.GDY".JAK.JAKI.JEŚLI.KTO.KTÓRY.
    PODCZAS.PONIEWAŻ.ŻE)
```

¹Pisane w marcu 2003 r.

\$ RÓŻNE (wf, BOK)
\$ RÓŻNE (t, ROZ).

KWER1 (wf, a, PRZY, t, rl, o, wa, wb, wc, k, i, AŻ1)
= FORMACZAS (wf, a, PRZY, t, rl, o, wa, wb, wc, k, i, AŻ1)
\$ RÓŻNE (wf, BOK) (WE33)
\$ RÓŻNE (t, ROZ).

KWER1 (wf, DK, c, t, rl, o, wa, wb, wc, k, i, z)
= FORMACZAS (wf, DK, c, t, rl, o, wa, wb, wc, k, i, z)
\$ RÓWNE (z, ZANIM.ZANIM'') (WE34)
\$ RÓŻNE (wf, BOK)
\$ RÓŻNE (t, ROZ).

KWER1 (wf, a, PRZE, OZN, rl, o, wa, wb, wc, k, i, z)
= FORMACZAS (wf, a, PRZE, OZN, rl, o, wa, wb, wc, k, i, z)
(WE35)
\$ RÓWNE (z, BY".CHOĆBY.CZYŻBY.GDYBY.JAKBY.JAKOBY.ŻEBY).

Dużymi literami w gramatyce zapisywane są stałe: nazwy jednostek składniowych jak **KWER1** (konstrukcja werbalna właściwa) i **FORMACZAS** (forma czasownikowa), nazwy wartości kategorii gramatycznych jak **PRZY** (czas przyszły) oraz poprzedzone znakiem dolara nazwy warunków. Małymi literami zapisywane są zmienne oznaczające nieokreślone wartości kategorii gramatycznych.

Konstrukcja werbalna właściwa jest więc opisana przez 6 reguł — oznaczonych symbolami od **WE30** do **WE35** — o podobnej treści: konstrukcja ta jest po prostu formą czasownikową, ale pod warunkiem zachodzenia pewnych relacji między wartościami kategorii gramatycznych stanowiących własności wymienionych jednostek składniowych; własności te nazywamy technicznie ich parametrami, a nazwy jednostek składniowych — nieterminalnymi symbolami gramatyki.

5 Kategorie gramatyczne konstrukcji werbalnej

Omówimy obecnie parametry symbolu nieterminalnego **KWER1**, czyli kategorie gramatyczne stanowiące własności konstrukcji werbalnej właściwej. Są one takie same, jak dla dla formy czasownikowej (**FORMACZAS**). Jest to szczególny przypadek ogólnej zasady, że konstrukcji równoważnej dystrybucyjnie pewnej formie wyrazowej przypisujemy własności posiadane przez tę formę.

Parametry te będziemy klasyfikować na wewnętrzne i zewnętrzne. Parametr wewnętrzny to taki, którego wartość wynika jednoznacznie z *formy* danego wyrazu czy konstrukcji. Dla wielu parametrów wewnętrznych ich wartość można łatwo ustalić również dla wyrazów nowych lub nieznanymi; jeśli nie jest to możliwe, to parametr taki nazywamy słownikowym. Parametry, które nie są wewnętrzne, nazywamy zewnętrznymi.

wf Wyróżnik fleksyjny. Parametr ten wskazuje, czy dana konstrukcja jest równoważna formie osobowej (**OS**), bezosobnikowi (**BOS**), bezokolicznikowi (**BOK**), imiesłowowi przysłówkowe-

mu uprzedniemu (PSU) czy imiesłowowi przysłówkowemu współczesnemu (PSW). Warto zauważyć, że wartość wyróżnika fleksyjnego wynika jednoznacznie z *formy* formy czasownikowej i w konsekwencji z *formy* konstrukcji werbalnej właściwej — jest to więc własność *wewnętrzna* tych jednostek składniowych.

a Aspekt. Jest to kategoria tradycyjna o dwóch wartościach — dokonany (DK) i niedokonany (ND); dla czasowników dwuaspektowych w niektórych konstrukcjach może nie mieć ustalonej wartości. Jest to również własność wewnętrzna.

c Czas. Jest to kategoria o trzech tradycyjnych wartościach — przeszły (PRZE), przyszły (PRZY) i teraźniejszy, dla którego Świdziński nie zaproponował żadnego symbolu, a który oznaczamy TER; również własność wewnętrzna.

Oczywiście, dla pewnych form czasownika nie jest ona w ogóle określona. Stosowany formalizm dopuszcza taką możliwość, zamiast konkretnej wartości występuje wówczas tzw. zmienna wolna, w wynikach analizatora AS reprezentowana jako liczba poprzedzona znakiem podkreślenia lub samo podkreślenie (np. 1923). Ta sama uwaga stosuje się również do innych kategorii wymienionych niżej.

t Tryb. Jest to także kategoria tradycyjna o trzech wartościach — oznajmujący (OZN), rozkazujący (ROZ) i warunkowy (WAR); również własność wewnętrzna.

rl Rodzaj i liczba. Świdziński potraktował rodzaj i liczbę jako jeden parametr, twierdząc bez szczegółowego uzasadnienia, że „poważnie upraszcza to opis” ([12], p. 4.6.2, s. 86). Z naszego punktu widzenia wystarczającym powodem do traktowania tych własności oddzielnie jest fakt, że liczba jest niekontrowersyjną własnością wewnętrzną, podczas gdy w przypadku kategorii rodzaju sprawa jest znacznie bardziej skomplikowana. Dla niektórych form i konstrukcji werbalnych, np. $\langle\langle$ pisze $\rangle\rangle$, następuje całkowita neutralizacja tej kategorii, co można interpretować dwojako: jako wieloznaczność lub „nieuchwytność” (ang. *vagueness*). Interpretacja pierwsza prowadzi do rosnącej wykładniczo liczby różnych rozbiorów gramatycznych — zdanie $\langle\langle$ Czeka, aż przyjdzie. $\rangle\rangle$ trzeba interpretować jako $\langle\langle$ On czeka, aż on przyjdzie. $\rangle\rangle$, $\langle\langle$ On czeka, aż ona przyjdzie. $\rangle\rangle$ itd. Podejście drugie wydaje się o wiele bardziej naturalne, wymaga to jednak rozszerzenia formalizmu o sposób reprezentacji nieookreślonych wartości kategorii rodzaju. Nie jest to zadanie trudne, ale ze względu na brak miejsca nie będziemy się nim tutaj zajmować.

o Osoba. Jest to kategoria tradycyjna o trzech wartościach, oznaczanych po prostu cyframi 1, 2 i 3; własność wewnętrzna.

wa, wb, wc Wymagania składniowe. Świdziński zakłada, że każdy czasownik wymaga — łącznie z podmiotem — co najwyżej trzech fraz odpowiedniego typu. Jest to własność słownikowa, nie można jej bezpośrednio wywnioskować z postaci formy czasownikowej. Wartość wymagania czasownikowego ma postać jednego z takich symboli, jak np. DK (faza werbalna bezokolicznikowa o aspekcie dokonanym), BEZ (faza przyimkowa z przyimkiem BEZ), MIAN (faza nominalna w mianowniku), MIAN' (faza przymiotnikowa w mianowniku), PRZYŚŁ (faza przysłówkowa), CZY (faza zdaniowa typu CZY). Jak widać, w stosunku do tradycyjnego pojęcia rządu czasownikowego repertuar wymagań czasownikowych został znacznie wzbogacony. Z drugiej strony repertuar ten ogranicza się do prostych symboli,

nie wykorzystując w pełni możliwości stosowanego formalizmu, który pozwala tutaj na użycie również symboli złożonych, co zostało wykorzystane w analizatorze AS.

- k** Korelatywność. Jest to nietradycyjna kategoria wprowadzona przez Świdzińskiego, która jest również własnością słownikową.

Genezą tej kategorii jest niekontrowersyjny fakt, że niektóre czasowniki, wymagające frazy zdaniowej, wymagają lub dopuszczają, aby była ona poprzedzona elementem nazywanym przez Klemensiewicza zapowiednikiem zespolenia, a przez Polańskiego — korelatem; por. [12], p. 8.1.4, s. 269. Oto dwa przykłady (s. 271):

(15) (a) Mówiono o tym, że grzmi.

(16) (a) Mówiono, że grzmi.

W pierwszym przykładzie wartością korelatywności jest 02, co w przyjętej przez Świdzińskiego konwencji oznacza przyimek *O* wymagający miejscownika; w drugim wartością jest NK (wartość niekorelatywna). Widać przy okazji, że występująca dwukrotnie w przykładach forma $\langle\langle Mówiono \rangle\rangle$ jest traktowana jako wystąpienia dwóch różnych choć homonimicznych czasowników — czasownika *MÓWIĆ* wymagającego korelatu i czasownika *MÓWIĆ* bez tego wymagania; dopuszczenie wymagań fakultatywnych pozwoliłoby łatwo uniknąć tej komplikacji.

Przyjęty przez Świdzińskiego sposób opisu tego typu konstrukcji budzi różnorodne wątpliwości. Po pierwsze, można łatwo uniknąć stosowania arbitralnych konwencji typu 01, 02 stosując przejrzyste symbole złożone, np. *O/BIER* i *O/MIEJ*. Po drugie, kategoria korelatywności jest przypisana wszystkim czasownikom i opartych na nich konstrukcjach werbalnych, co nie wydaje się niczym uzasadnione. Po trzecie, nawet w przypadku czasowników wymagających frazy zdaniowej, własność ta odnosi się nie tyle do samego czasownika, co do jego konkretnego wymagania. Wydaje się więc, że optymalnym sposobem opisu tego zjawiska jest poszerzenie repertuaru wymagań składniowych o symbole złożone zawierające w sobie zarówno informację o typie wymaganej frazy zdaniowej (np. *ZE*), jak i o jej korelacji (np. *O/MIEJ*). Sposób zapisu tej informacji nie jest w gruncie rzeczy istotny, może to być np. *O/MIEJ+ZE* lub inny podobny symbol.

Trzeba jednak podkreślić, że powyższe niedogodności nie mają istotnego wpływu na przebieg i efektywność automatycznej analizy syntaktycznej.

- i** Inkorporacja. Nazwa tej kategorii pochodzi od spójników inkorporacyjnych czyli nieinicjalnych (*BOWIEM*, *NATOMIAST*, *WIĘC*, *ZAŚ*). Wartością tej kategorii jest albo nazwa odpowiedniego spójnika, albo symbol NI (wartość nieinkorporacyjna). Wartość w postaci nazwy spójnika oznacza, że w danej konstrukcji składniowej wystąpił właśnie ten spójnik.

Świdziński wyraźnie się wahał, czy kategorię inkorporacji przypisać formie czasownikowej — w oryginalnej wersji reguł *WE30*, *WE31*, *WE32* symbol *FORMACZAS* występuje bez parametru inkorporacji, a z parametrem w regułach *WE33*, *WE34* i *WE35*. W wykazie jednostek składniowych na s. 330 symbol *FORMACZAS* występuje również bez parametru inkorporacji, i w konsekwencji tak to zostało przyjęte w analizatorze AS.

Trzeba jednak tutaj podkreślić, że forma czasownikowa (*FORMACZAS*) — jako oczywista — jest niezdefiniowana w gramatyce Świdzińskiego; niezbędne reguły musiały zostać dodane na potrzeby analizatora AS. Warto także przypomnieć, że gramatyka Świdzińskiego

jest szczególnym przypadkiem tzw. gramatyki struktur frazowych, która w języku angielskim nazywa się bardziej intuicyjnie *immediate constituent grammar* czyli gramatyką składników bezpośrednich. W konsekwencji symbol FORMACZAS może oznaczać tylko jakąś spójną sekwencję wyrazów. Reguły dodane w analizatorze AS pozwalają zatem na jego realizację przez takie formy złożone czasownika, których człony występują obok siebie. Aby w przyszłości akceptować również takie konstrukcje, jak $\langle\langle$ będzie natomiast czytał $\rangle\rangle$, należało przyjąć, że forma czasownikowa posiada jednak kategorię inkorporacji, która stanowi jej własność wewnętrzną, bo w oczywisty sposób jej wartość wynika z postaci formy czasownikowej.

- z Zależność. Ta nietradycyjna kategoria gramatyczna jest chyba najbardziej charakterystyczną cechą gramatyki Świdzińskiego. Stanowi ona z jednej strony o jej sile wyrazu i precyzji, z drugiej strony przyczynia się istotnie do tego, że gramatyka ta jest mało przejrzysta i trudna do ogarnięcia. Wartością zależności jest jeden z 40 symboli — takich jak P (wartość pytajna), NP (wartość niepytająca), NP' 't (wartość niepytająca z ograniczeniem czasu, trybu, aspektu i negacji).

Pełne omówienie kategorii zależności daleko wykracza poza ramy niniejszego artykułu, zajmijmy się więc dalej tylko pewnymi jej aspektami.

Przede wszystkim należy zwrócić uwagę, że wartość tej kategorii nie wynika z formy konstrukcji werbalnej, ale z kontekstu jej użycia. Jest to więc przykład *własności zewnętrznej*. Występując w warunkach reguł od WE30 do WE35 decyduje ona o tym, czy dana konstrukcja werbalna właściwa będzie uznana za poprawną w danym kontekście jej użycia.

Popatrzmy teraz na omówione kategorie gramatyczne z nieco innego punktu widzenia.

Wyróżnik fleksyjny, aspekt, czas, tryb, liczba, inkorporacja, częściowo rodzaj, to wartości wewnętrzne formy czasownikowej, dające się — przynajmniej w zasadzie — jednoznacznie ustalić na podstawie jej kształtu. Konstrukcja werbalna właściwa ma identyczne wartości tych kategorii, jak będąca jej jedynym składnikiem forma czasownikowa. Wymagania składniowe i korelatywność to własności, które dla danej formy czasownikowej muszą być pobrane z odpowiedniego słownika gramatycznego. Ich wspólną cechą z wymienionymi wcześniej kategoriami jest to, że również ich wartości są identyczne dla formy czasownikowej i zawierającej ją konstrukcji werbalnej właściwej. W takich sytuacjach wygodnie jest mówić, że konstrukcja werbalna *propaguje* — czyli przekazuje bez zmian — odpowiednie własności formy czasownikowej.

W przypadku kategorii zależności mamy sytuację dokładnie odwrotną. To forma czasownikowa ma identyczną wartość tej kategorii, jak konstrukcja werbalna właściwa, ta druga zaś otrzymuje tę wartość z reguły opisującej kontekst jej wystąpienia. W takim wypadku wygodnie jest mówić, że forma czasownikowa *dziedziczy* własność zależności po konstrukcji werbalnej właściwej, ta zaś po jej konstrukcji nadrzędnej.

Jak wspomnieliśmy, kategoria zależności decyduje o poprawności danej konstrukcji werbalnej w danym kontekście. Podobnie jak matematycy zastanawiają się czasami, jak pewne założenie *interweniuje* w dowodzie pewnego twierdzenia, tak my możemy stawiać sobie pytania, jak pewna kategoria lub jej konkretna wartość *interweniuje* w uznaniu poprawności pewnej konstrukcji składniowej. W przypadku kategorii zależności i konstrukcji werbalnej właściwej odpowiedź jest oczywista — wartość kategorii zależności interweniuje w warunkach reguł od WE30

do WE35. Możemy sobie stawiać jednak bardziej subtelne pytania, np. czy ta forma interwencji kategorii zależności wydaje się optymalna.

Można w szczególności postawić pytanie, czy słusznie gramatyka Świdzińskiego przypisuje kategorię zależności również formie czasownikowej; własność tę forma czasownikowa po prostu dziedziczy po nadrzędnej konstrukcji werbalnej, ma ona sens czysto informacyjny, bo nie jest na tym poziomie w żaden sposób wykorzystywana. W konsekwencji nie ma ona żadnego wpływu na przebieg analizy syntaktycznej i dlatego w gramatyce analizatora AS została ona usunięta.

6 Wartość AŻ1 kategorii zależności

Kategoria zależności jest między innymi wykorzystywana do opisu zjawiska omawianego przez Annę Łojasiewicz w książce [7]. Na stronach 72-73 dyskutuje ona dwa użycia spójnika AŻ, pisząc m.in.: *Spójnik aż w jednym z dwu swych użyciach rozróżnionych w tej pracy, a mianowicie w pozycji konotowanej wymaga, aby czasownik w członie przez niego wprowadzanym miał formę czasu przyszłego.*

Chodzi tu o wystąpienia konotowane przez czasownik CZEKAĆ i pochodne (DOCZEKAĆ, ZACZEKAĆ itd.). Ilustrują to m.in. następujące przykłady.

Czekaj, aż przyjdę.
*Czekaj, aż przyszedłam.
*Czekaj, aż przychodzę.

Niekonotowane użycia spójnika AŻ, kiedy wprowadza on frazę luźną, pozbawione tego ograniczenia, ilustrują m.in. następujące przykłady.

Krzyczy, aż uszy bołą
Łże, aż się kurzy.
Nalegał, aż ustąpili.

W odniesieniu do konotowanego użycia spójnika AŻ Łojasiewicz stwierdza również: *wydaje się tu istnieć jakieś ograniczenie dotyczące aspektu.* Pisze ona dalej: *Nasuwa się następujące przypuszczenie. Niezręczne jest użycie czasownika niedokonanego po spójniku AŻ, jeśli istnieje dla tego czasownika odpowiednik dokonany.*

Referując te poglądy Świdziński pisze ([12], punkt 8.6.2.1, s. 295–296) *A. Łojasiewicz zwraca uwagę na istnienie różnych użyci jednostki AŻ we współczesnej polszczyźnie (Łojasiewicz (1983: 94-96, 145)). Wyróżnia ona użycia AŻ w kontekście czasownika CZEKAĆ, gdzie — według autorki — niemożliwa jest niedokonaność składnika zdaniowego pospójnikowego..*

Świdziński także rozróżnia dwa typy użyci, wprowadzając dwa spójniki: AŻ1 i AŻ2. Pisze on dalej: *Fraza zdaniowa typu AŻ1 wykazuje mocne ograniczenie czasu. Musi ona mieć mianowicie wartość czasu przyszłego [...]* *A. Łojasiewicz nie ma racji dopatrując się w konstrukcjach z AŻ1 ograniczenia aspektu.* Aby uzasadnić to ostatnie stwierdzenie, Świdziński podaje poprawne jego zdaniem przykłady, zawierające spójnik AŻ w obu jego użyciach (s. 296)²:

²Jeden z recenzentów tomu, prof. Maciej Grochowski, napisał m.in., że nie rozumie tych przykładów. Ja też ich nie rozumiem, służą one tutaj tylko do zilustrowania poglądów autora gramatyki.

- (118) Będą czekać, aż przyjdziecie, aż przyjdziecie.
 (119) Poczekajcie, aż ich zaprosimy, aż ich nie zaprosimy.
 (120) Aż majster da mi znak, zaczekać warto, aż przyjdziecie.

Warto tutaj zwrócić uwagę na następujące sprawy. Po pierwsze, nie wydaje się, aby poglądy Łojasiewicz w pracy Świdzińskiego były przedstawione dostatecznie jasno i wiernie. Po drugie, ani Łojasiewicz, ani Świdziński nie interesują się ograniczeniami czasu we frazie luźnej, np.

Krzyczy, aż uszy bolą.
 *Krzyczy, aż uszy bolały.

Najważniejszy jest jednak dla mnie fakt, że kiedy Łojasiewicz mówi o dwóch użyciach spójnika, Świdziński mówi o dwóch spójnikach. Jak się wydaje, jest to całkowicie niepotrzebna i nieuzasadniona komplikacja opisu.

7 Przykład analizy

Popatrzmy obecnie na wynik analizy syntaktycznej zdania

Czekaliśmy, aż przyjdziecie.

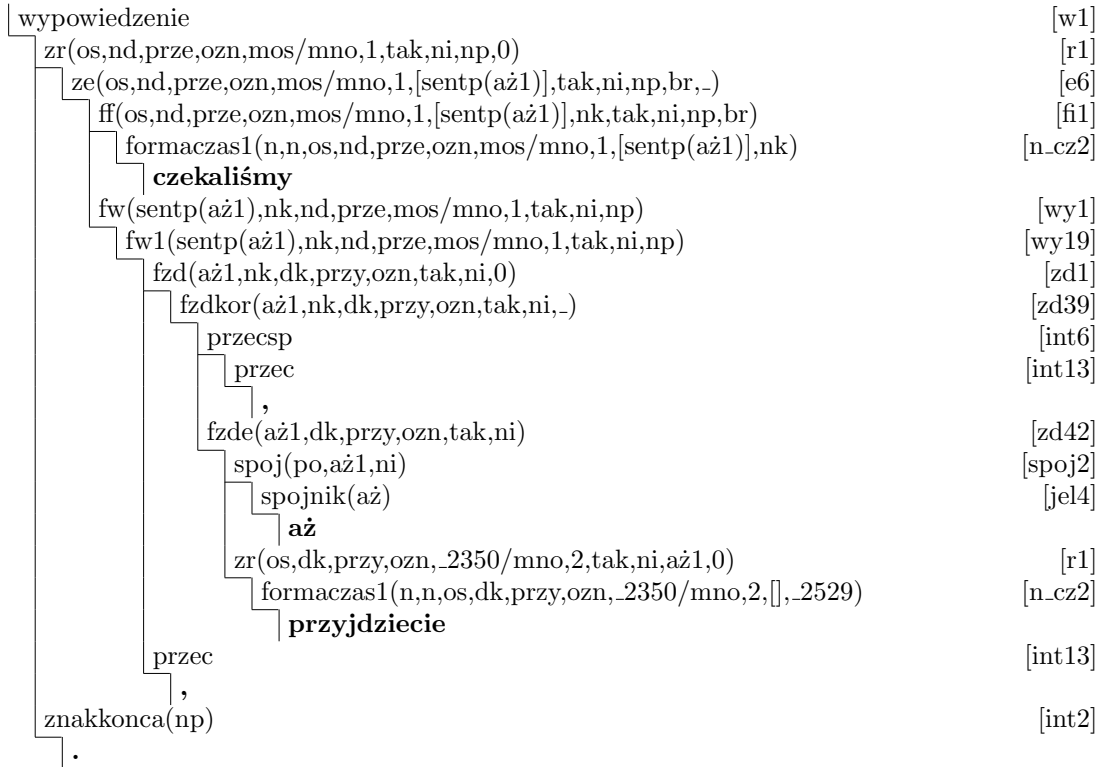
Jest to przykład (147) znajdujący się na s. 200 w rozdziale 6 książki [12]. Rysunek 1 przedstawia strukturę drzewa rozbioru, a rysunek 2 — pełne drzewo. W analizatorze AS i jego wynikach stosowana jest nieco inna notacja dla kategorii gramatycznych i ich własności. W szczególności zamiast trzech wymagań składniowych stosowana jest lista wymagań, ujęta w nawiasy prostokątne. Więcej informacji o tych konwencjach notacyjnych — w dużym stopniu przejętych z wcześniejszego analizatora AMOS — można znaleźć we wspomnianych pracach [1], [2] [3].

Jak można zauważyć na rys. 2 we wierszu oznaczonym [we33], konstrukcja werbalna właściwa zbudowana z formy czasownikowej *przyjdziecie* ma — co nie jest dla nas zaskoczeniem — wartość kategorii zależności postaci AŻ1.

Wartość ta jest bez żadnych zmian odziedziczona po składowym zdaniu równorzędnym ZR, za pośrednictwem między innymi takich konstrukcji składniowych, jak fraza finitywna (FF), fraza werbalna (FWE) czy konstrukcja werbalna z inkorporacją (KWEINK). Zdanie równorzędne otrzymuje tę wartość dzięki regule ZD42, która — po poprawieniu błędu literowego — ma postać następującą:

$$\begin{aligned} \text{FZDE} (\text{tfz}, \text{a}, \text{c}, \text{t}, \text{neg}, \text{i}) \\ = \text{SPÓJ} (\text{PO}, \text{tfz}, \text{i}) & \qquad \qquad \qquad (\text{ZD42}) \\ \text{ZR} (\text{wf}, \text{a}, \text{c}, \text{t}, \text{rl}, \text{o}, \text{neg}, \text{NI}, \text{tfz}) \\ \$ \text{RÓWNE} (\text{tfz}, \text{AŻ1}.\text{AŻ2}.\text{CHOCIAŻ}.\text{CZY}.\text{DOPÓKI}.\text{GDY}.\text{JAK}.\text{JEŚLI}.\text{PODZAS}.\text{PONIEWAŻ}.\text{ZANIM}.\text{ŻE}) \end{aligned}$$

Efekt użycia tej reguły widać na obydwu rysunkach — są to fragmenty drzewa zaczynające się we wierszu oznaczonym [zd42]. W szczególności widać tam wyraźnie, że w regule tej następuje jakby zmiana statusu symbolu AŻ1. Jest on z jednej strony wartością kategorii *typ frazy zdaniowej*, stanowiącej własność frazy zdaniowej elementarnej (FZDE), z drugiej strony — jak



Rysunek 1: Struktura drzewa analizy

już wiemy — jest to wartość kategorii zależności zdania równorzędnego. Fraza zdaniowa elementarna dziedziczy wartość typu frazy zdaniowej po frazie wymaganej (FW) za pośrednictwem innych konstrukcji, takich jak fraza zdaniowa szeregowa (FZD) i fraza zdaniowa z korelatem (FZDKOR). Ze względów technicznych we frazie wymaganej i frazie wymaganej właściwej (FW1) symbol ten ma bardziej złożoną postać.

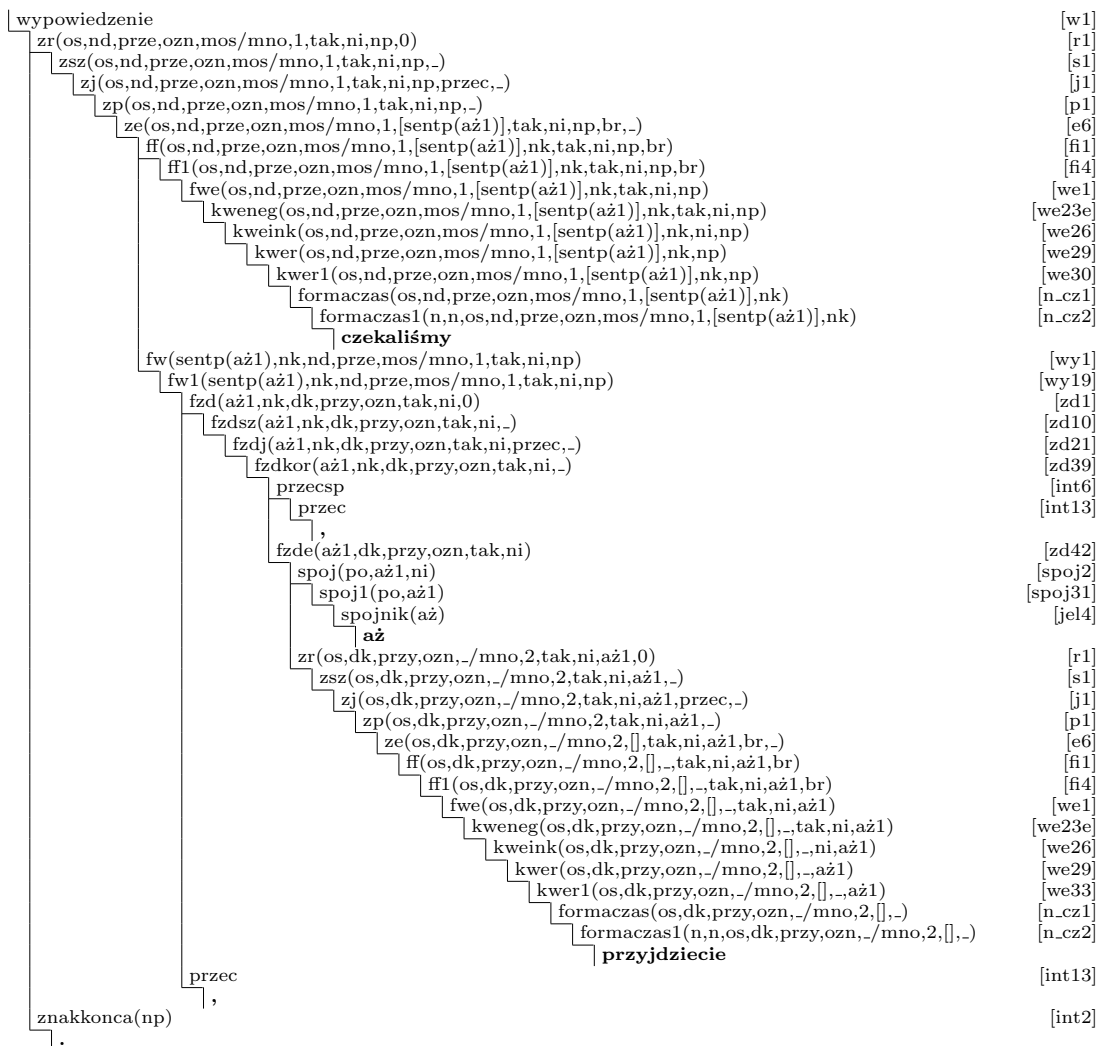
Podobna zmiana statusu symbolu AŻ1 dokonuje się w regule E6:

```

ZE (wf, a, c, t, r1, o, wa, wb, wc, neg, i, z, ow)
= FF (wf, a, c, t, r1, o, wa, wb, wc, k, neg, i, z, ow)
  FW (wa, k, a, c, r1, o, neg, NI, NP) (E6)
  FW (wb, k, a, c, r1, o, neg, NI, NP)
  FW (wc, k, a, c, r1, o, neg, NI, NP)
  $ RÓŻNE (z, BY".CHOĆBY.CO.CZYŻBY.GDYBY.JAKBY.JAKI.JAKOBY.
    KTO.KTÓRY.P.P'.P".PZ.ŻEBY).

```

Efekt użycia tej reguły również widać na obydwu rysunkach — są to fragmenty drzewa zaczynające się we wierszu oznaczonym [e6]. Na rysunkach nie są jednak zaznaczone te gałęzie drzewa, które odpowiadają pustym realizacjom fraz wymaganych.



Rysunek 2: Pełne drzewo analizy

W regule tej wymagania frazy finitywnej (FF) utożsamiane są z typami fraz wymaganych odpowiednich fraz (FW). Wymaganie AŻ1 pochodzi oczywiście — za pośrednictwem odpowiednich konstrukcji składniowych — od formy czasownikowej *czekaliśmy*, dla której jest to własność słownikowa. Warto dodać, że w regule tej występuje również inne użycie kategorii zależności, które nie dotyczy obecnych rozważań (w naszym przykładzie to wystąpienie kategorii zależności ma wartość NP czyli *niepytajne*).

Znając drogę, jaką przechodzi symbol AŻ1 od słownika czasowników do jego wystąpienia jako wartości zależności w konstrukcji werbalnej właściwej opartej na wyrazie $\langle\langle\text{przyjdziecie}\rangle\rangle$, można zbadać, jaki ma on wpływ na przebieg analizy. Wartość ta jest wykorzystywana w warunkach kilku reguł. Część z nich już przytaczaliśmy wyżej, oto pozostałe:

FF1 (wf, a, c, t, rl, o, wa, wb, wc, k, neg, i, z, ow)
 = FWE (wf, a, c, t, rl, o, wa, wb, wc, k, neg, i, z1) (FI4)
 \$ RÓWNE (wf, OS.BOS.BOK)
 \$ OBLZAL (z1, z, ow).

KWENEG (wf, a, c, t, rl, o, wa, wb, wc, k, TAK, i, z)
 = KWEINK (wf, a, c, t, rl, o, wa, wb, wc, k, i, z) (WE23)
 FL (a, c, rl, o, TAK, NI, NP)
 \$ RÓŻNE (z, AŻ1".BO".BY".DOPÓKI.GDY".P.P'.P".PZ.ZANIM").

W regule FI14 wartość zależności frazy werbalnej (FWE) nie jest dziedziczona bezpośrednio po frazie finitywnej właściwej (FF1), lecz jest obliczana jako wypadkowa wartości zależności tej frazy (z) oraz wartości innej nietradycyjnej kategorii występującej w gramatyce Świdzińskiego, a mianowicie *ograniczenia wewnętrznego* (ow). W pewnych kontekstach wartość zależności AŻ1 może zostać przekształcona na AŻ1'', nie będziemy się jednak tą sprawą tutaj zajmować.

W regule WE23 wartość kategorii zależności konstrukcji werbalnej z negacją (KWENEG) jest istotna tylko dla frazy luźnej, więc nie dotyczy to naszego przykładu.

W regule E6 praktycznym efektem tego, że wartość AŻ1 spełnia podany warunek, jest nadanie frazom wymaganim (FW) jako wartości kategorii zależności symbolu NP (wartość niepytajna), co dla naszego przykładu też nie ma większego znaczenia.

Najbardziej istotną rolę wartość ta odgrywa w bloku reguł dla konstrukcji werbalnej właściwej, gdzie powoduje, że jedyną możliwą do zastosowania regułą jest reguła WE33. Reguła ta może być jednak zastosowana tylko wtedy, gdy forma czasownikowa jest w czasie przyszłym (parametr czasu symbolu FORMACZAS ma ustaloną wartość PRZY), oraz gdy spełnione są dwa dodatkowe, w gruncie rzeczy redundantne warunki: forma czasownikowa nie może być bezokolicznikiem (zatem wyróżnik fleksyjny wf musi być różny od BOK) i jej tryb nie może być rozkazujący (czyli parametr t musi być różny od ROZ). W ten właśnie sposób jest egzekwowane wymaganie czasu przyszłego we frazach konotowanych przez czasowniki takie, jak CZEKAĆ.

8 Podsumowanie

Sformułowanie formalnego opisu składni języka naturalnego wymaga podejmowania licznych decyzji, często konwencjonalnych, czasami czysto arbitralnych, niekiedy narzuconych przez stosowany formalizm. Z takimi decyzjami mamy do czynienia również w gramatyce Świdzińskiego i omawianym przykładzie. Z jednej strony mamy pewną klasę czasowników, która nakłada

pewne wymagania na konotowaną frazę zdaniową. Z drugiej strony mamy formę czasownikową frazy konotowanej, która ma pewne określone własności. Gramatyka formalna musi umożliwić skonfrontowanie konkretnych wymagań z konkretnymi własnościami. Aby to było możliwe, odpowiednia informacja musi być przekazywana w drzewie analizy syntaktycznej za pomocą odpowiednich parametrów. Forma tych parametrów nie jest już wyznaczona przez naturę opisywanego zjawiska, lecz ustalona arbitralnie przez autora gramatyki. Arbitralny jest również wybór reguły gramatycznej, w której następuje konfrontacja wymagań z własnościami. W naszym przykładzie jest to reguła WE33, ale nic nie stoi na przeszkodzie, aby warunki tej reguły przenieść np. do odpowiednio rozbudowanej reguły E6. Powstaje zatem pytanie, które rozwiązanie jest lepsze, przy czym rozwiązania te można porównywać i oceniać na podstawie różnych, może nawet sprzecznych kryteriów: efektywność komputerowej analizy, przejrzystość, modyfikowalność i rozszerzalność gramatyki formalnej, przejrzystość i intuicyjność drzew analizy itd. Dzięki analizatorom AS i Świgrą poszukiwanie odpowiedzi na pytania tego typu stało się istotnie łatwiejsze.

Literatura cytowana

- [1] Janusz S. Bień. Komputerowa weryfikacja opisu składni polskiej. Raport Instytutu Informatyki Uniwersytetu Warszawskiego TR 96-06 (227), maj 1996. <ftp://ftp.mimuw.edu.pl/pub/users/polszczyzna/AMOS-95/tr227.ps> 3, 7
- [2] Janusz S. Bień. Komputerowa weryfikacja formalnej gramatyki Świdzińskiego. *Biuletyn Polskiego Towarzystwa Językoznawczego* zeszyt LII (1997), s. 147-164. 3, 7
- [3] Janusz Stanisław Bień. Zestaw testów do weryfikacji i oceny analizatorów języka polskiego. Sprawozdanie merytoryczne (nieznacznie zmodyfikowane) z projektu KBN 8 T11C 002 13. <ftp://ftp.mimuw.edu.pl/pub/users/polszczyzna/tajp/>. 3, 7
- [4] Janusz S. Bień, Krzysztof Szafran, Marcin Woliński. Experimental parsers of Polish. 3. *Europäische Konferenz "Formale Beschreibung slavischer Sprachen, Leipzig 1999"*. Linguistische Arbeitsberichte 75, Institut für Linguistik, Universität Leipzig, 2000, pp 185-190. December 2000. 3
- [5] Colmerauer, A. 1978. Metamorphosis grammar. In: L. Bolc (ed.), *Natural Language Communication with Computers*, Lecture Notes in Computer Science 63, Springer-Verlag 1978, pp 133-189. 2
- [6] Feliks Kluźniak, Stanisław Szpakowicz. *Prolog*. WNT: Warszawa 1983. 2
- [7] Anna Łojasiewicz. *Własności składniowe polskich spójników*. Wydawnictwa UW 1992. 6
- [8] Krzysztof Szafran. Analizator morfologiczny SAM-95 — opis użytkowy. Raport Instytutu Informatyki Uniwersytetu Warszawskiego TR 96-05 (226), maj 1996. <ftp://ftp.mimuw.edu.pl/pub/users/polszczyzna/SAM-95/tr226.ps>. 3
- [9] Krzysztof Szafran, Automatyczne hasłowanie tekstu polskiego. *Polonica* tom XVIII. IJP PAN: Kraków 1997, s. 51-63. 3

- [10] Stanisław Szpakowicz, *Formalny opis składniowy zdań polskich*, wydanie drugie, Wydawnictwa UW: Warszawa 1986. Wersja elektroniczna dostępna w Internecie <ftp://ftp.mimuw.edu.pl/pub/user/polszczyzna/Szpakowicz/>. 2
- [11] Stanisław Szpakowicz, Marek Świdziński, 1990. *Formalna definicja równorzędnej grupy nominalnej we współczesnej polszczyźnie pisanej*. Studia Gramatyczne IX (1990), s. 9-54. 2
- [12] Marek Świdziński, *Gramatyka formalna języka polskiego*. Wydawnictwa Uniwersytetu Warszawskiego 1992. 2, 4, 5, 6, 7
- [13] Jan Tokarski. *Schematyczny indeks a tergo polskich form wyrazowych*. Opracowanie i redakcja Zygmunt Saloni. Wydanie drugie. Wydawnictwo Naukowe PWN, Warszawa 2002. 3