

AGNIESZKA PATEJUK*

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA

ADAM PRZEPIÓRKOWSKI**

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK, WARSZAWA; UNIWERSYTET WARSZAWSKI

POLFIE: współczesna gramatyka formalna języka polskiego

Słowa kluczowe: gramatyka, LFG, polski, składnia.

1. Wstęp

Niniejszy artykuł przedstawia nową gramatykę formalną języka polskiego – POLFIE, która jest rozwijana od 2010 roku w ramach projektów realizowanych w Zespole Inżynierii Lingwistycznej IPI PAN (m.in. NEKST¹ i CLARIN-PL²). POLFIE łączy prace w nurcie teoretycznym – gramatyka opiera się na jednej z najważniejszych teorii lingwistycznych, oraz implementacyjnym – gramatyka stanowi podstawę automatycznego systemu analizy składniowej tekstów polskich. Artykuł zawiera krótki opis teorii lingwistycznej leżącej u podstaw gramatyki (§2) oraz prezentację wybranych zjawisk lingwistycznych uwzględnionych w gramatyce POLFIE (§3), a także wspomina o komputerowej implementacji i dostępności gramatyki (§4).

2. Teoria LFG

Lexical Functional Grammar (LFG; Bresnan (red.) 1982; Dalrymple 2001; Bresnan i in. 2015) jest teorią lingwistyczną aktywnie rozwijaną od drugiej połowy lat siedemdziesiątych XX wieku. Chociaż LFG jest teorią generatywną w tym sensie, że ma solidne podstawy formalne, nie należy do chomskiańskich gramatyk transformacyjnych. LFG jest też teorią silnie zleksykalizowaną – wiele zjawisk składniowych można opisać bezpośrednio w leksykonie (zamiast w ogólnych regułach składniowych), co często pozwala na subtelniejszy opis.

Teoria LFG oferuje bogatą reprezentację dzięki zastosowaniu wielu równoległych, powiązanych ze sobą poziomów reprezentacji – dwa podstawowe poziomy to c-struktura i f-struktura. C-struktura to struktura składnikowa (ang. *c-structure* od *constituent structure*) w formie drzewa, powstała na podstawie kategorii składniowych, takich jak fraza nominalna (NP) czy fraza czasownikowa (IP), natomiast f-struktura to struktura funkcyjna (ang. *f-structure* od *functional structure*) w formie struktury atrybutów (ang. *attribute-value matrix*, AVM),

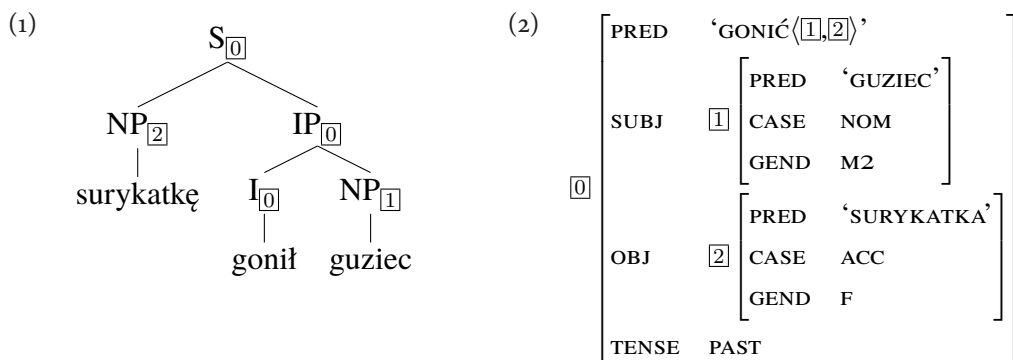
* agnieszka.patejuk@ipipan.waw.pl

** adam.przepiorkowski@ipipan.waw.pl

¹ <http://zil.ipipan.waw.pl/NEKST>.

² <http://clip.ipipan.waw.pl/CLARIN-PL>.

powstała na podstawie takich funkcji gramatycznych jak podmiot (SUBJ) czy dopełnienie bliższe (OBJ), ale zawierająca też inne atrybuty. Przykładowe struktury dla zdania *Surykatkę gonił guziec* podane są poniżej. Powiązania między strukturami oddane są za pomocą zmiennych takich jak [1], która w tym wypadku ustala związek pomiędzy frazą nominalną (NP) *guziec* w c-strukturze w (1) i podmiotem (SUBJ) w f-strukturze w (2).



Funkcje gramatyczne są podstawowym pojęciem teorii LFG – w przeciwieństwie do teorii chomskiańskich funkcje gramatyczne nie są pochodną pozycji w drzewie, co jest istotne dla języków o swobodnym szyku, takich jak język polski. Funkcje gramatyczne nie są również bezpośrednio zależne od kategorii składniowych: podmiot nie musi być nominalny, ponieważ znane jest zjawisko podmiotów niekanonicznych, np. zdaniowych (*Jana dziwiło, że Maria wybiera Piotra*), które mogą być koordynowane z podmiotami nominalnymi (*Jana dziwiło, że Maria wybiera Piotra, i jej brak gustu*; por. Świdziński 1992b, 1993). Funkcje gramatyczne nie są też zależne od roli semantycznej, co jest widoczne na przykładzie pasywizacji, gdzie dopełnienie strony czynnej (*Surykatkę gonił guziec*) staje się podmiotem strony biernej (*Surykatka była goniona (przez guźca)*) mimo zachowania tej samej roli semantycznej. Funkcje gramatyczne są zatem pojęciem składniowym, które określa relację między nadrzędnikiem i jego podrzędnikiem – mogą być w pewnym stopniu zależne od pozycji w drzewie (szczególnie w językach z ustalonym szykiem), kategorii składniowej czy roli semantycznej, ale nie mogą być do żadnego z tych pojęć sprowadzone.

Nadrzędnik w relacji opisanej daną funkcją gramatyczną reprezentowany jest za pomocą atrybutu PRED. Wartość tego złożonego atrybutu zawiera semantyczną charakterystykę nadrzędnika, na którą składają się: wyrażony tym nadrzędnikiem predykat (reprezentowany jako jego lemat, np. GONIC) oraz argumenty tego predykatu, zwykle podane jako zmienne odwołujące się do wartości odpowiednich funkcji gramatycznych (zob. [1] i [2] w (2)). Argumenty predykatu dzielą się na dwa rodzaje: argumenty semantyczne, które są elementami listy zamkniętej w nawiasach kątowych (<...>), i ewentualne argumenty niesemantyczne (zob. §3.2 poniżej), które podane są poza tą listą (zob. [1] w (10) i [2] w (11)). Istnieją też predykaty, których lista argumentów jest pusta: np. rzeczowniki GUZIEC i SURYKATKA w (2) – ich atrybut PRED ma wartość składającą się wyłącznie z lematu.

Poza funkcjami gramatycznymi i atrybutem PRED f-struktura zawiera również inne atrybuty, m.in. morfoskładniowe, takie jak CASE i GEND w (2)–(3), które określają wartość przypadku i rodzaju gramatycznego³. Możliwe jest wprowadzenie dowolnych atrybutów, które są potrzebne do analizy, np. atrybutu NEG, który umożliwia opisanie zależności przypadku dopełnienia bliższego od negacji: w zdaniu *Guziec nie gonił surykatki*, któremu odpowiada f-struktura w (3), obecność negacji (NEG) powoduje wystąpienie dopełniacza negacji, natomiast w zdaniu bez negacji, *Guziec gonił surykatkę*, z f-strukturą jak w (2), wymagany jest biernik.

(3)

[0]	PRED	'GONIC'([1],[2])'							
	SUBJ	[1]	<table> <tr> <td>PRED</td> <td>'GUZIEC'</td> </tr> <tr> <td>CASE</td> <td>NOM</td> </tr> <tr> <td>GEND</td> <td>M2</td> </tr> </table>	PRED	'GUZIEC'	CASE	NOM	GEND	M2
			PRED	'GUZIEC'					
	CASE	NOM							
	GEND	M2							
OBJ	[2]	<table> <tr> <td>PRED</td> <td>'SURYKATKA'</td> </tr> <tr> <td>CASE</td> <td>GEN</td> </tr> <tr> <td>GEND</td> <td>F</td> </tr> </table>	PRED	'SURYKATKA'	CASE	GEN	GEND	F	
		PRED	'SURYKATKA'						
		CASE	GEN						
GEND	F								
NEG	+								
TENSE	PAST								

3. Analizy wybranych zjawisk

Niniejszy punkt przedstawia zwięzłe analizy niektórych zjawisk objętych gramatyką POLFIE; pełniejsze opisy znaleźć można w literaturze cytowanej w poszczególnych podpunktach.

3.1. Kontrola i podnoszenie

Formalizm LFG umożliwia reprezentację znanych z literatury generatywnej zjawisk kontroli składniowej i podnoszenia⁴. Niektóre predykaty wymagają bezokolicznika (*Antek kazał/obiecał Erykowi zaśpiewać*), którego podmiot nie może być wyrażony leksykalnie, natomiast jego rozumienie jest ściśle określone przez predykat wymagający bezokolicznika – w wypadku czasownika KAZAĆ jest to jego argument celownikowy (*Erykowi*), natomiast w wypadku OBIECAĆ jest to jego podmiot (*Antek*). Teoria LFG obejmuje takie zjawiska opisem, dając możliwość określenia takich zależności między argumentami predykatów (argumentem czasownika głównego a podmiotem bezokolicznika) za pomocą relacji kontroli, która zapisywana jest bezpośrednio w słowniku, we wpisach leksykalnych odpowiednich predykatów: wpis dla KAZAĆ w (4) wskazuje dopełnienie dalsze (OBJ_θ) jako podmiot bezokolicznika

3 Dla oszczędności miejsca pominięte zostały inne atrybuty morfoskładniowe, w tym NUM oznaczający liczbę gramatyczną.

4 Ze stosunkowo niedawnych prac wspomnieć można monografię Landau 2013, a także – w ujęciu minimalistycznym (Chomsky 1995) i z odniesieniem do polszczyzny – prace: Witkoś i in. 2011ab.

(XCOMP SUBJ), natomiast wpis dla OBIECAĆ wskazuje podmiot (SUBJ) jako podmiot bezokolicznika (XCOMP SUBJ).

$$(4) \quad (\uparrow \text{PRED}) = \text{'KAZAĆ'} < (\uparrow \text{SUBJ}) (\uparrow \text{OBJ}_\theta) (\uparrow \text{XCOMP}) >$$

$$(\uparrow \text{OBJ}_\theta) = (\uparrow \text{XCOMP SUBJ})$$

$$(5) \quad (\uparrow \text{PRED}) = \text{'OBIECAĆ'} < (\uparrow \text{SUBJ}) (\uparrow \text{OBJ}_\theta) (\uparrow \text{XCOMP}) >$$

$$(\uparrow \text{SUBJ}) = (\uparrow \text{XCOMP SUBJ})$$

Różnica w kontroli między predykatami KAZAĆ i OBIECAĆ wynikająca z odpowiednich wpisów w leksykonie jest odzwierciedlona w f-strukturach: (6) odpowiada zdaniu *Antek kazał Erykowi zaśpiewać*, natomiast (7) jest reprezentacją zdania *Antek obiecał Erykowi zaśpiewać*.

$$(6) \quad \left[\begin{array}{ll} \text{PRED} & \text{'KAZAĆ'} \langle \boxed{1}, \boxed{2}, \boxed{3} \rangle \\ \text{SUBJ} & \boxed{1} \left[\begin{array}{ll} \text{PRED} & \text{'ANTEK'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{array} \right] \\ \text{OBJ}_\theta & \boxed{2} \left[\begin{array}{ll} \text{PRED} & \text{'ERYK'} \\ \text{CASE} & \text{DAT} \\ \text{GEND} & \text{M1} \end{array} \right] \\ \text{XCOMP} & \boxed{3} \left[\begin{array}{ll} \text{PRED} & \text{'ZAŚPIEWAĆ'} \langle \boxed{2} \rangle \\ \text{SUBJ} & \boxed{2} \end{array} \right] \\ \text{TENSE} & \text{PAST} \end{array} \right]$$

$$(7) \quad \left[\begin{array}{ll} \text{PRED} & \text{'OBIECAĆ'} \langle \boxed{1}, \boxed{2}, \boxed{3} \rangle \\ \text{SUBJ} & \boxed{1} \left[\begin{array}{ll} \text{PRED} & \text{'ANTEK'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{array} \right] \\ \text{OBJ}_\theta & \boxed{2} \left[\begin{array}{ll} \text{PRED} & \text{'ERYK'} \\ \text{CASE} & \text{DAT} \\ \text{GEND} & \text{M1} \end{array} \right] \\ \text{XCOMP} & \boxed{3} \left[\begin{array}{ll} \text{PRED} & \text{'ZAŚPIEWAĆ'} \langle \boxed{1} \rangle \\ \text{SUBJ} & \boxed{1} \end{array} \right] \\ \text{TENSE} & \text{PAST} \end{array} \right]$$

Jak wspomniano w §2, w f-strukturach do oznaczenia argumentów danego predykatu użyte są zmienne; w obu omawianych przykładach zmienna $\boxed{1}$ odnosi się do podmiotu czasownika osobowego (*Antek*), zaś $\boxed{2}$ – do argumentu celownikowego (*Erykowi*) analizowanego tutaj jako dopełnienie dalsze (OBJ_θ). Trzecim argumentem obydwu predykatów jest sygnalizowany zmienną $\boxed{3}$ argument bezokolicznikowy, XCOMP. W przeciwieństwie do innych omawianych dotychczas typów argumentów XCOMP należy do argumentów otwartych – jego podmiot musi być kontrolowany przez argument wyższego predykatu i nie może być zrealizowany leksykalnie⁵. W f-strukturze (6) dla KAZAĆ podmiotem predykatu ZAŚPIEWAĆ jest $\boxed{2}$, czyli argument celownikowy predykatu KAZAĆ; natomiast w f-strukturze (7) dla OBIECAĆ podmiotem predykatu ZAŚPIEWAĆ jest $\boxed{1}$ – podmiot predykatu OBIECAĆ.

Należy podkreślić, że reprezentacja kontroli składniowej ma nie tylko znaczenie semantyczne; następny podpunkt pokazuje, że ma również wpływ na takie zjawiska składniowe jak uzgodnienie.

5 Relacja kontroli zawarta we wpisie leksykalnym danego czasownika (np. $(\uparrow \text{OBJ}_\theta) = (\uparrow \text{XCOMP SUBJ})$ w (4)) określa, który argument czasownika kontroli jest tożsamy z podmiotem kontrolowanego bezokolicznika – z tego powodu nie jest możliwe wypełnienie pozycji podmiotu kontrolowanego bezokolicznika w inny sposób.

3.2. Argumenty predykatywne

Teoria LFG umożliwia również adekwatną reprezentację argumentów predykatywnych w zdaniach takich jak *Marysia wydaje się Antkowi miła* czy *Marysia uważa Antka za miłego* – w obu argumentem predykatywnym jest forma przymiotnika MIŁY (w drugim zdaniu zagnieżdżona we frazie przyimkowej), ale odnosi się ona do różnych argumentów czasowników: podmiotu w pierwszym zdaniu i dopełnienia bliższego w drugim. Reprezentacja tych zależności możliwa jest przy użyciu mechanizmu kontroli opisanego w §3.1 – odpowiednie wpisy w leksykonie wskazują, do czego odnosi się argument predykatywny (XCOMP-PRED): w wypadku WYDAWAĆ SIĘ wpis leksykalny (8) wskazuje, że argument predykatywny odnosi się do podmiotu (SUBJ), natomiast według wpisu (9) dla czasownika UWAŻAĆ argument taki odnosi się do dopełnienia bliższego (OBJ).

- (8) $(\uparrow \text{PRED}) = \text{'WYDAWAĆ SIĘ'} < (\uparrow \text{OBJ}_\theta) (\uparrow \text{XCOMP}) > (\uparrow \text{SUBJ})'$
 $(\uparrow \text{SUBJ}) = (\uparrow \text{XCOMP-PRED SUBJ})$
- (9) $(\uparrow \text{PRED}) = \text{'UWAŻAĆ'} < (\uparrow \text{SUBJ}) (\uparrow \text{XCOMP}) > (\uparrow \text{OBJ})'$
 $(\uparrow \text{OBJ}) = (\uparrow \text{XCOMP-PRED SUBJ})$

W wyniku powyższych specyfikacji powstają następujące f-struktury: (10) odpowiada zdaniu *Marysia wydaje się Antkowi miła*, natomiast (11) reprezentuje zdanie *Marysia uważa Antka za miłego*. Specyfikacja w (8) sprawia, że w (10) wartość funkcji SUBJ dopełnienia predykatywnego XCOMP-PRED jest tożsąca z wartością funkcji SUBJ czasownika WYDAWAĆ SIĘ, co zasygnalizowane jest wielokrotnym wystąpieniem zmiennej [1]. Z kolei specyfikacja w (9) sprawia, że w (11) podmiotem dopełnienia predykatywnego jest dopełnienie bliższe czasownika UWAŻAĆ, co zasygnalizowane jest za pomocą zmiennej [2].

- (10)
$$\left[\begin{array}{cc} \text{PRED} & \text{'WYDAWAĆ SIĘ'} \langle [2], [3] \rangle [1] \\ \text{SUBJ} & [1] \left[\begin{array}{cc} \text{PRED} & \text{'MARYSIA'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{F} \end{array} \right] \\ \text{OBJ}_\theta & [2] \left[\begin{array}{cc} \text{PRED} & \text{'ANTEK'} \\ \text{CASE} & \text{DAT} \\ \text{GEND} & \text{M1} \end{array} \right] \\ \text{XCOMP-PRED} & [3] \left[\begin{array}{cc} \text{PRED} & \text{'MIŁY'} \langle [1] \rangle \\ \text{SUBJ} & [1] \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{F} \end{array} \right] \\ \text{TENSE} & \text{PRES} \end{array} \right]$$
- (11)
$$\left[\begin{array}{cc} \text{PRED} & \text{'UWAŻAĆ'} \langle [1], [3] \rangle [2] \\ \text{SUBJ} & [1] \left[\begin{array}{cc} \text{PRED} & \text{'MARYSIA'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{F} \end{array} \right] \\ \text{OBJ} & [2] \left[\begin{array}{cc} \text{PRED} & \text{'ANTEK'} \\ \text{CASE} & \text{ACC} \\ \text{GEND} & \text{M1} \end{array} \right] \\ \text{XCOMP-PRED} & [3] \left[\begin{array}{cc} \text{PRED} & \text{'MIŁY'} \langle [2] \rangle \\ \text{SUBJ} & [2] \\ \text{CASE} & \text{ACC} \\ \text{GEND} & \text{M1} \\ \text{PFORM} & \text{ZA} \end{array} \right] \\ \text{TENSE} & \text{PRES} \end{array} \right]$$

Analiza, według której elementy predykatywne mają podmiot (SUBJ), wynika m.in. z tego, że w niektórych językach, w tym w języku polskim czy rosyjskim, mogą one wystąpić bez łącznika, np. *Antek miły*. W takim wypadku – ponieważ w teoriach typu LFG z zasady unika się wprowadzania elementów, które nie występują na powierzchni (w rozważanym przykładzie byłby to niewyrażony łącznik) – powstaje f-struktura w (12), w której nadrzędnym predykatem jest przymiotnik predykatywny MIŁY⁶.

$$(12) \quad \left[\begin{array}{ll} \text{PRED} & \text{'MIŁY } \langle \boxed{1} \rangle' \\ & \left[\begin{array}{ll} \text{PRED} & \text{'ANTEK'} \\ \text{SUBJ } \boxed{1} & \left[\begin{array}{ll} \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{array} \right] \end{array} \right] \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{array} \right]$$

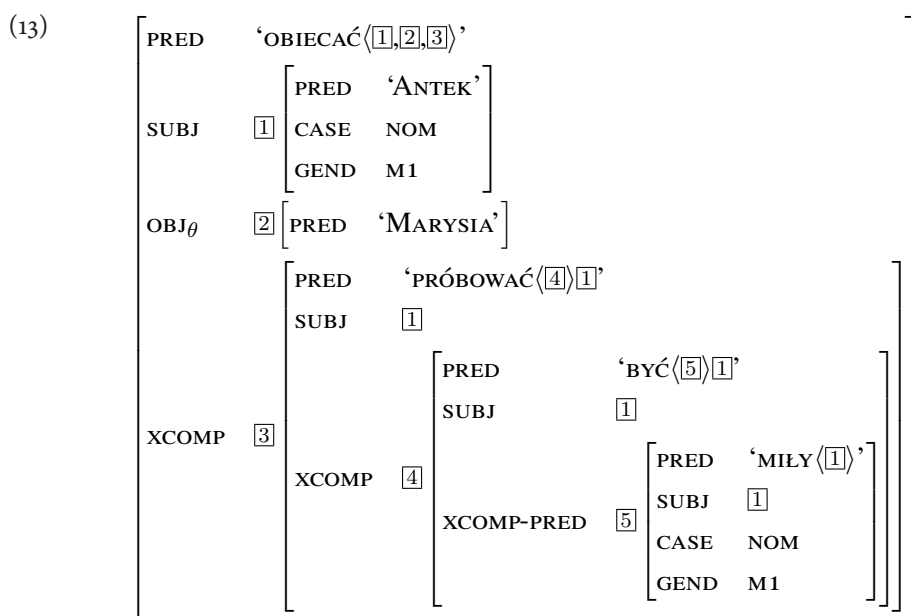
Jawne wskazanie elementu, do którego odnosi się dopełnienie predykatywne, pozwala na zapewnienie odpowiedniego uzgodnienia między tymi elementami – rodzaj i liczba przymiotnika predykatywnego muszą się uzgadniać z odpowiadającymi atrybutami argumentu, do którego ten przymiotnik się odnosi (por. niegramatyczność zdania **Marysia wydaje się Antkowi miły* czy też **Marysia uważa Antka za miłych*). Zapewnienie takiego uzgodnienia wymaga odpowiedniej reprezentacji zależności między tymi argumentami – teoria LFG to umożliwia.

Inaczej jest z przypadkiem gramatycznym przymiotnika predykatywnego: podczas gdy w wypadku WYDAWAĆ SIĘ podlega on uzgodnieniu z podmiotem czasownika, przy UWAŻAĆ jest on narzucany przez przyimek ZA, a w wypadku PREDYKATU UDAWAĆ jest to tzw. przypadek strukturalny, czyli taki, którego wartość morfologiczna zależy nie tylko od nadrzędnego predykatu, ale także od innych cech otoczenia składniowego, w tym od negacji (Patejuk, Przepiórkowski 2014c; *Marysia udaje miłą, Marysia nie udaje miłej*).

F-struktura dla zdania *Antek obiecał Marysi próbować być miły* w (13) pokazuje, jak łatwo można poradzić sobie z dalekim uzgodnieniem predykatywnym w formalizmie LFG. W zdaniu tym *Antek* jest podmiotem czasownika OBIECAĆ, którego podrzędnikiem bezokolicznikowym jest forma bezokolicznikowa czasownika PRÓBOWAĆ, który z kolei ma własne dopełnienie bezokolicznikowe w postaci formy czasownika BYĆ wymagającego dopełnienia predykatywnego (forma przymiotnika MIŁY). Zatem element, z którym uzgadnia się dopełnienie predykatywne, *Antek*, i samo dopełnienie predykatywne, *miły*, znajdują się na różnych poziomach struktury składniowej. Opisanie takiego dalekiego uzgodnienia predykatywnego wymaga odpowiedniej reprezentacji składniowej, którą formalizm LFG oferuje – w wyniku relacji kontroli składniowej podmiot OBIECAĆ przypisany do zmiennej $\boxed{1}$ jest współdzielony z podmiotem PRÓBOWAĆ, a następnie z podmiotem BYĆ – ten ostatni predykat współdzieli swój

6 Omówienie innych możliwych analiz takich konstrukcji znaleźć można w artykule Patejuk, Przepiórkowski 2014b.

podmiot z dopełnieniem predykatywnym (XCOMP-PRED). W rezultacie w f-strukturze argumentu predykatywnego informacja o atrybutach podmiotu predykatu OBIECAĆ (przypadek, rodzaj, liczba) dostępna jest lokalnie, dzięki czemu możliwe⁷ jest zapewnienie uzgodnienia z elementem predykatywnym. W ten sposób zjawisko pozornie nielokalne sprowadzone jest do sekwencji trzech lokalnych oddziaływań składniowych: dwa razy kontroli i raz uzgodnienia argumentu predykatywnego z podmiotem.



Szczegółową analizę takich i bardziej skomplikowanych konstrukcji z argumentami predykatywnymi w języku polskim znaleźć można w artykułach: Patejuk, Przepiórkowski 2014b oraz Przepiórkowski, Patejuk 2012b.

3.3. Haploglogia się

Teoria LFG umożliwia również adekwatne opisanie zjawiska haploglogii się w języku polskim. Istnieją predykaty, które wymagają wystąpienia elementu się, np. *Antek boi *(się)*⁸ czy *Antek śmieje *(się)*; w obu wypadkach występuje tzw. się inherentne (bez kontrybucji semantycznej). Możliwe jednak jest opuszczenie się, jeżeli występuje ono już przy wyższym strukturalnie czasowniku, np. *Antek bał się śmiać*. Co więcej, w łańcuchu czasowników (ciągu następujących po sobie bezokoliczników) między predykatami wymagającymi się może wystąpić predykat, z którym się nie występuje, np. w zdaniu *Antek bał się próbować śmiać*.

7 Podczas gdy w omawianym przykładzie występuje uzgodnienie przypadku, nie jest ono możliwe, kiedy kontrolę sprawuje dopełnienie, jak np. w zdaniu *Antek kazał Marysi próbować być miłą/*miał* – zamiast uzgodnienia wymagany jest tzw. narzędnik predykcji (opisany w POLFIE), możliwy również w przykładzie z OBIECAĆ (*Antek obiecał Marysi próbować być miłym*).

8 Zapis taki oznacza, że zdanie jest gramatyczne, gdy element w nawiasie występuje, ale niegramatyczne bez tego elementu; oznaczająca niegramatyczność gwiazdka odnosi się więc do opcjonalności sygnalizowanej nawiasami.

F-struktura w (14) odpowiada obu wersjom zdania *Antek bał się śmiać (się)* (tj. z obecnym lub opuszczonym drugim *się*), natomiast (15) odpowiada obu wersjom zdania *Antek bał się próbować śmiać (się)*. Teoria LFG umożliwia odnoszenie się do wyższych i niższych warstw f-struktury: za pomocą ograniczenia w (16) możliwe jest sprawdzenie, czy *SIĘ* występuje lokalnie (jak w *Antek śmieje się*) lub w wyższym predykanie w łańcuchu czasowników (jak w *Antek boi się (próbować) śmiać*). Ograniczenie to działa następująco: operator $*$ przy funkcji *xcomp* odpowiadającej dopełnieniu bezokolicznikowemu oznacza dowolnie wiele (w tym zero) wystąpień tej funkcji po kolei – zatem $(xcomp^* \uparrow)$ pozwala wejść dowolnie daleko w górę, do wyższej f-struktury w łańcuchu czasowników (z *ŚMIAĆ SIĘ* do *PRÓBOWAĆ* i dalej do *BAĆ SIĘ*). Następnie sprawdzane ($=_c$) jest, czy wartość atrybutu *SIE* w tym fragmencie f-struktury to $+$, co oznacza wystąpienie *SIĘ*.

$$(14) \left[\begin{array}{ll} \text{PRED} & \text{'BAĆ SIĘ} \langle \boxed{1}, \boxed{2} \rangle \\ \text{SUBJ} & \boxed{1} \left[\begin{array}{ll} \text{PRED} & \text{'ANTEK'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{array} \right] \\ \text{XCOMP} & \boxed{2} \left[\begin{array}{ll} \text{PRED} & \text{'ŚMIAĆ SIĘ} \langle \boxed{1} \rangle \\ \text{SUBJ} & \boxed{1} \\ \text{SIE} & + \end{array} \right] \\ \text{SIE} & + \end{array} \right]$$

$$(15) \left[\begin{array}{ll} \text{PRED} & \text{'BAĆ SIĘ} \langle \boxed{1}, \boxed{2} \rangle \\ \text{SUBJ} & \boxed{1} \left[\begin{array}{ll} \text{PRED} & \text{'ANTEK'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \end{array} \right] \\ \text{XCOMP} & \boxed{2} \left[\begin{array}{ll} \text{PRED} & \text{'PRÓBOWAĆ} \langle \boxed{3} \rangle \boxed{1} \\ \text{SUBJ} & \boxed{1} \\ \text{XCOMP} & \boxed{3} \left[\begin{array}{ll} \text{PRED} & \text{'ŚMIAĆ SIĘ} \langle \boxed{1} \rangle \\ \text{SUBJ} & \boxed{1} \\ \text{SIE} & + \end{array} \right] \end{array} \right] \\ \text{SIE} & + \end{array} \right]$$

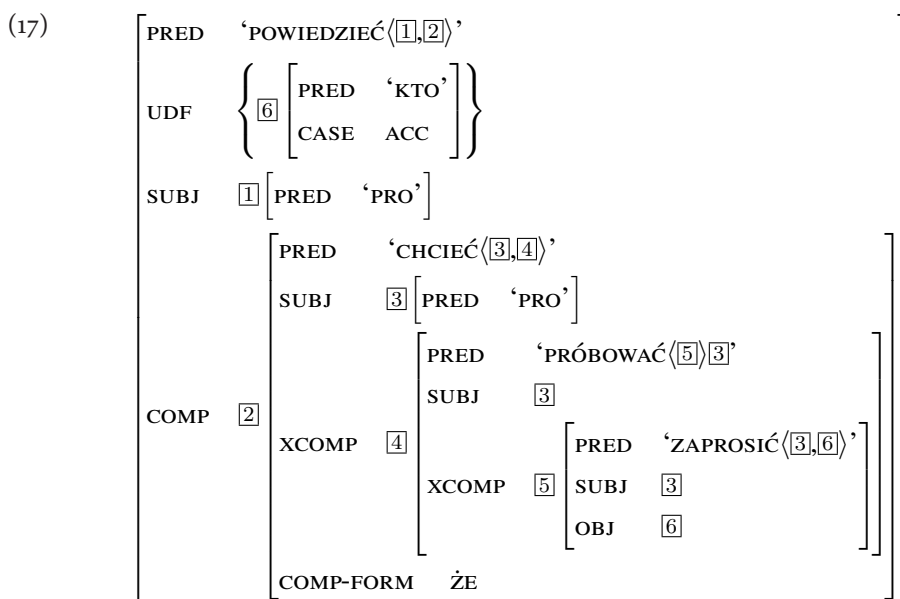
$$(16) ((xcomp^* \uparrow) \text{SIE}) =_c +$$

Przedstawiona powyżej analiza jest uproszczona, ponieważ nie odróżnia leksykalnego wystąpienia *SIĘ* (*Antek bał się śmiać się*) od haplogicznego (*Antek bał się śmiać*). Nie uwzględnia

też reprezentacji różnych typów SIĘ oraz wielofunkcyjności SIĘ, jak w jednym z odczytań zdania *Śmiało się z kłopotów*, gdzie SIĘ jest jednocześnie inherentne (wymagane przez czasownik ŚMIAĆ SIĘ) oraz bezosobowe (konstrukcyjne; por. *Dobrze czytało się ten artykuł*). Obydwa aspekty są objęte pełną analizą przedstawioną w artykule Patejuk i Przepiórkowski 2015a.

3.4. Zależności nielokalne

Teoria LFG umożliwia także opis zależności nielokalnych (ang. *unbounded dependencies*) – pod tym pojęciem kryje się m.in. zjawisko zwane czasami ekstrakcją (ang. *wh-extraction*), widoczne w takich zdaniach jak: *Kogo powiedziałeś, że chcesz próbować zaprosić? Kogo znajduje się tu w linearnym otoczeniu formy predykatu POWIEDZIEĆ, chociaż nie może być analizowane jako jego argument; jest natomiast argumentem predykatu ZAPROSIĆ, który jest dopełnieniem bezokolicznikowym (XCOMP) predykatu PRÓBOWAĆ, a ten z kolei jest dopełnieniem bezokolicznikowym predykatu CHCIEĆ, który w końcu jest dopełnieniem zdaniowym (COMP) predykatu POWIEDZIEĆ. Zatem zaimek Kogo znajduje się we frazie – odpowiadającej czasownikowi POWIEDZIEĆ – oddalonej o trzy poziomy zagnieżdżenia (COMP XCOMP XCOMP) od predykatu ZAPROSIĆ, który wymaga tego zaimka jako dopełnienia (OBJ) i warunkuje wartość jego przypadku gramatycznego. W omawianej gramatyce zdaniu takiemu odpowiada f-struktura w (17).*



Powyższa f-struktura jest wynikiem interakcji odpowiednich wpisów leksykalnych i reguł składniowych. W szczególności reguła składniowa odpowiedzialna za realizację zaimka pytającego przy czasowniku osobowym przypisuje temu zaimkowi specyfikację w (18); strzałka w dół, ↓, oznacza tu sam zaimek, sygnalizowany w (17) zmienną [6], zaś strzałka w górę, ↑, oznacza

f-strukturę zdania. Specyfikacja $\downarrow \in (\uparrow \text{UDF})$ powoduje zatem dodanie zaimka pytajnego do zbioru UDF, w którym znajdują się elementy podlegające ekstrakcji, zaś specyfikacja $\downarrow = (\uparrow \text{COMP XCOMP}^* \text{OBJ})$ powoduje, że *Kogo* jest jednocześnie dopełnieniem (OBJ) jednego z ciągu bezokoliczników (XCOMP*) w zagnieżdżonej frazie zdaniowej (COMP) – tutaj ciągu bezokoliczników *próbować zaprosić*. Zaimek pytajny jest zatem argumentem predykatu ZAPROSIĆ, który wpływa na przypadek gramatyczny tego zaimka – wobec braku negacji jest to biernik.

$$(18) \quad \begin{aligned} \downarrow &\in (\uparrow \text{UDF}) \\ \downarrow &= (\uparrow \text{COMP XCOMP}^* \text{OBJ}) \end{aligned}$$

3.5. Koordynacja

Formalizm LFG pozwala także na objęcie opisem rozmaitych zjawisk składniowych związanych z koordynacją – zostaną tu przedstawione tylko ciekawsze i trudniejsze zjawiska.

Podczas gdy łatwe jest opisanie współdzielenia podrzędnika przy koordynacji, gdy podrzędnik znajduje się na lewym lub prawym krańcu frazy skoordynowanej (*Antek szedł i śpiewał* czy *Kupił i od razu przeczytał dwie książki*), trudniejsze są przypadki, gdy podrzędnik jest „wpleciony”, tj. znajduje się wewnątrz jednej ze skoordynowanych fraz (*Szedł Antek i śpiewał*). W ramach teorii LFG możliwa jest analiza, według której współdzielony podrzędnik *Antek*, wpleciony w jeden z członów koordynacji, reprezentowany jest analogicznie do podrzędnika na krańcu frazy skoordynowanej (Patejuk i Przepiórkowski 2015a). Wspólna f-struktura dla obu powyższych zdań – *Antek szedł i śpiewał* oraz *Szedł Antek i śpiewał* – dana jest w (19), gdzie podmiot obu predykatów oznaczony jest zmienną $\boxed{1}$. Gramatyka natomiast oczywiście przypisze tym zdaniom różne c-struktury, co oddaje różnicę w szyku.

$$(19) \quad \left[\left[\begin{array}{cc} \text{PRED} & \text{'IŚĆ}(\boxed{1}) \\ \text{SUBJ} & \boxed{1} \end{array} \right] \left[\begin{array}{cc} \text{PRED} & \text{'ANTEK'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{M1} \\ \text{NUM} & \text{SG} \end{array} \right] \left[\begin{array}{cc} \text{PRED} & \text{'ŚPIEWAĆ}(\boxed{1}) \\ \text{SUBJ} & \boxed{1} \\ \text{TENSE} & \text{PAST} \end{array} \right] \right] \\ \text{COORD-FORM} \quad \text{I}$$

Teoria LFG umożliwia także sformalizowanie zjawiska koordynacji różnych kategorii (Przepiórkowski, Patejuk 2012a; Patejuk, Przepiórkowski 2014a; Patejuk 2015), w którego ramach koordynowane są frazy o odmiennej kategorii składniowej, odpowiadające jednak jednej pozycji składniowej, a zatem mające wspólną funkcję gramatyczną – f-struktura w (20) odpowiada zdaniu *Chcę pić i papierosa* (Kallas 1993: 123), gdzie dopełnienie (OBJ) jest frazą skoordynowaną, która składa się z bezokolicznika (*pić*) i rzeczownika (*papierosa*).

$$(20) \left[\begin{array}{l} \text{PRED} \quad \text{'CHCIEĆ'} \langle \boxed{1}, \boxed{2} \rangle \\ \text{SUBJ} \quad \boxed{1} \left[\text{PRED} \quad \text{'PRO'} \right] \\ \text{OBJ} \quad \boxed{2} \left\{ \begin{array}{l} \left[\begin{array}{l} \text{PRED} \quad \text{'PIĆ'} \langle \boxed{1} \rangle \\ \text{SUBJ} \quad \boxed{1} \\ \text{CONTROLLER} \quad \boxed{1} \end{array} \right], \left[\begin{array}{l} \text{PRED} \quad \text{'PAPIEROS'} \\ \text{CASE} \quad \text{ACC} \end{array} \right] \\ \text{COORD-FORM} \quad \text{I} \\ \text{CONTROLLER} \quad \boxed{1} \end{array} \right\} \end{array} \right]$$

Reprezentacja LFG oddaje fakt, że predykat CHCIEĆ jest czasownikiem wymagającym kontroli bezokolicznika przez podmiot (podmiot domyślny CHCIEĆ przypisany do zmiennej $\boxed{1}$ jest też podmiotem PIĆ), jednak kontrola dotyczy jedynie bezokolicznikowego członu koordynacji, a nie rzeczownika (*papierosa*). Ponadto wzięte jest pod uwagę nadawanie przypadku strukturalnego: dotyczy ono z kolei jedynie rzeczownikowego członu koordynacji (*papierosa*), natomiast nie ma zastosowania w wypadku członu bezokolicznikowego. Nadawanie przypadku jest lepiej widoczne w przykładzie, w którym nie występuje forma o biernikowo-dopełniaczowym synkretyzmie przypadku: *Chcę pić i kanapkę*, gdzie bez negacji KANAPKA występuje w bierniku, natomiast w obecności negacji, *Nie chcę pić ani kanapki*, występuje w dopełniaczu (przykład analogiczny do *Nie chciał pić ani kanapki*; Kallas 1993: 92).

Ostatnim wspomnianym tutaj interesującym zjawiskiem dotyczącym koordynacji, które zostało uwzględnione w omawianej gramatyce, a może przysparzać problemów innym teoriom, jest koordynacja różnych funkcji gramatycznych (Patejuk, Przepiórkowski 2012; Przepiórkowski, Patejuk 2014; Patejuk 2015), znana również pod nazwą koordynacji leksykalno-semantycznej. Jest ona możliwa pod warunkiem, że poszczególne elementy koordynacji należą do tego samego typu semantycznego (np. zaimki pytajne, negatywne, upowszechniające, typu *-kolwiek* itd.) – jak w zdaniu *Jakie i skąd zdobywał informacje?* (Kallas 1993: 141), gdzie skoordynowane są zaimki pytajne. W tym przykładzie koordynacja łączy formę *Jakie* – przymiotnik będący modyfikatorem rzeczownika *informacje*, który jest dopełnieniem bliższym (OBJ) czasownika – i formę *skąd*, która jest z kolei modyfikatorem samego czasownika. Skoordynowane są tu zatem frazy pochodzące z różnych poziomów f-struktury (modyfikator czasownika i modyfikator dopełnienia tego czasownika).

Teoria LFG umożliwia analizę formalną tego zjawiska i stworzenie f-struktury, która oddaje zarówno fakt koordynacji danych elementów, jak i to, że odpowiadają one innym funkcjom gramatycznym, a ponadto mogą być podrzędnikami innych nadrzędników – reprezentacją omawianego zdania jest f-struktura w (21). Zaproponowana analiza umieszcza skoordynowane elementy w atrybucie UDF i, analogicznie jak przy ekstrakcji omówionej w §3.4, współdzieli je z odpowiednimi elementami f-struktury: zaimek przymiotny *Jakie* jest przypisany do zmiennej $\boxed{3}$ i odpowiada modyfikatorowi (ADJ) dopełnienia (OBJ) formy czasownika ZDOBYWAĆ, a więc modyfikatorowi rzeczownika *informacje*, natomiast zaimek przysłówkowy *skąd*

przypisany jest do zmiennej [4] i odpowiada modyfikatorowi (ADJ) czasownika. Taka analiza pozwala na zapewnienie właściwego uzgodnienia między dopełnieniem (*informacje*) a jego modyfikatorem (*Jakie*) mimo tego, że są od siebie znacznie oddalone linearnie. Pełną analizę tego zjawiska przedstawia rozprawa doktorska Patejuk 2015.

$$(21) \left[\begin{array}{l} \text{PRED} \quad \text{'ZDOBYWAĆ'} \langle [1], [2] \rangle \\ \text{SUBJ} \quad [1] \left[\begin{array}{l} \text{PRED} \quad \text{'PRO'} \end{array} \right] \\ \text{OBJ} \quad [2] \left[\begin{array}{l} \text{PRED} \quad \text{'INFORMACJA'} \\ \text{CASE} \quad \text{ACC} \\ \text{ADJ} \quad \{ [3] \} \end{array} \right] \\ \text{ADJ} \quad \{ [4] \} \\ \text{UDF} \quad \left\{ \left[\begin{array}{l} [3] \left[\begin{array}{l} \text{PRED} \quad \text{'JAKI'} \\ \text{CASE} \quad \text{ACC} \\ \text{TYPE} \quad \text{INT} \end{array} \right], [4] \left[\begin{array}{l} \text{PRED} \quad \text{'SKĄD'} \\ \text{TYPE} \quad \text{INT} \end{array} \right] \right] \right\} \\ \text{COORD-FORM} \quad \text{I} \end{array} \right]$$

4. Implementacja i dostępność

LFG jest na tyle sformalizowaną teorią lingwistyczną, że możliwe jest zapisanie gramatyki w postaci umożliwiającej automatyczną analizę składniową za pomocą dedykowanej tej teorii platformy implementacyjnej XLE stworzonej przez Xerox PARC⁹. Istnienie opartego na gramatyce analizatora składniowego, tzw. parsera, ma fundamentalne znaczenie dla rozwoju gramatyki, istotnie ułatwia bowiem weryfikację przyjętych w gramatyce rozwiązań i umożliwia prześledzenie interakcji pomiędzy modułami gramatyki opisującymi różne, ale czasem współwystępujące zjawiska. Prezentowana tu gramatyka POLFIE została w pełni zaimplementowana w środowisku XLE. Co więcej, gramatyka jest rozwijana w ramach międzynarodowego projektu PARGRAM (Parallel Grammar Project; Butt i in. 2002; <http://pargram.b.uib.no/>), czego efektem jest nacisk na stosowanie w gramatyce mechanizmów i analiz, o których wiadomo, że są wspólne dla większej liczby języków, a nie arbitralnych rozwiązań, działających w wypadku polszczyzny, ale niekoniecznie dających się zastosować do opisu podobnych zjawisk w innych językach. Gramatyka POLFIE jest członkiem projektu PARGRAM od roku 2011 i jako jedyna reprezentuje w projekcie języki słowiańskie; rozwijane są również gramatyki następujących języków: angielskiego, greckiego (nowożytnego), indonezyjskiego, niemieckiego, norweskiego (bokmål i nynorsk), urdu, węgierskiego i wolof.

9 Xerox Linguistic Environment (<http://www2.parc.com/isl/groups/nlt/xle/>).

Implementacja gramatyki POLFIE maksymalizuje korzystanie z istniejących zasobów językowych dla języka polskiego: gramatyka powstała na bazie dwóch wcześniejszych zaimplementowanych gramatyk języka polskiego, a mianowicie GFJP (Gramatyki Formalnej Języka Polskiego; gramatyki DCG wykorzystywanej przez parser Świgr; Świdziński 1992a; Woliński 2004), która stała się inspiracją dla reguł tworzących drzewa POLFIE, oraz FOJP (Formalnego Opisu Języka Polskiego; gramatyki HPSG towarzyszącej publikacji: Przepiórkowski i in. 2002), która była inspiracją dla struktur funkcyjnych POLFIE. Ponadto POLFIE korzysta z dwóch zasobów do tworzenia leksykonu: z analizatora Morfeusz2 (Woliński 2014) jako źródła informacji morfoskładniowej i ze słownika Walenty (Przepiórkowski i in. 2014, 2017; Hajnicz i in. 2016) jako źródła informacji o wymaganiach walencyjnych predykatów (Patejuk 2016).

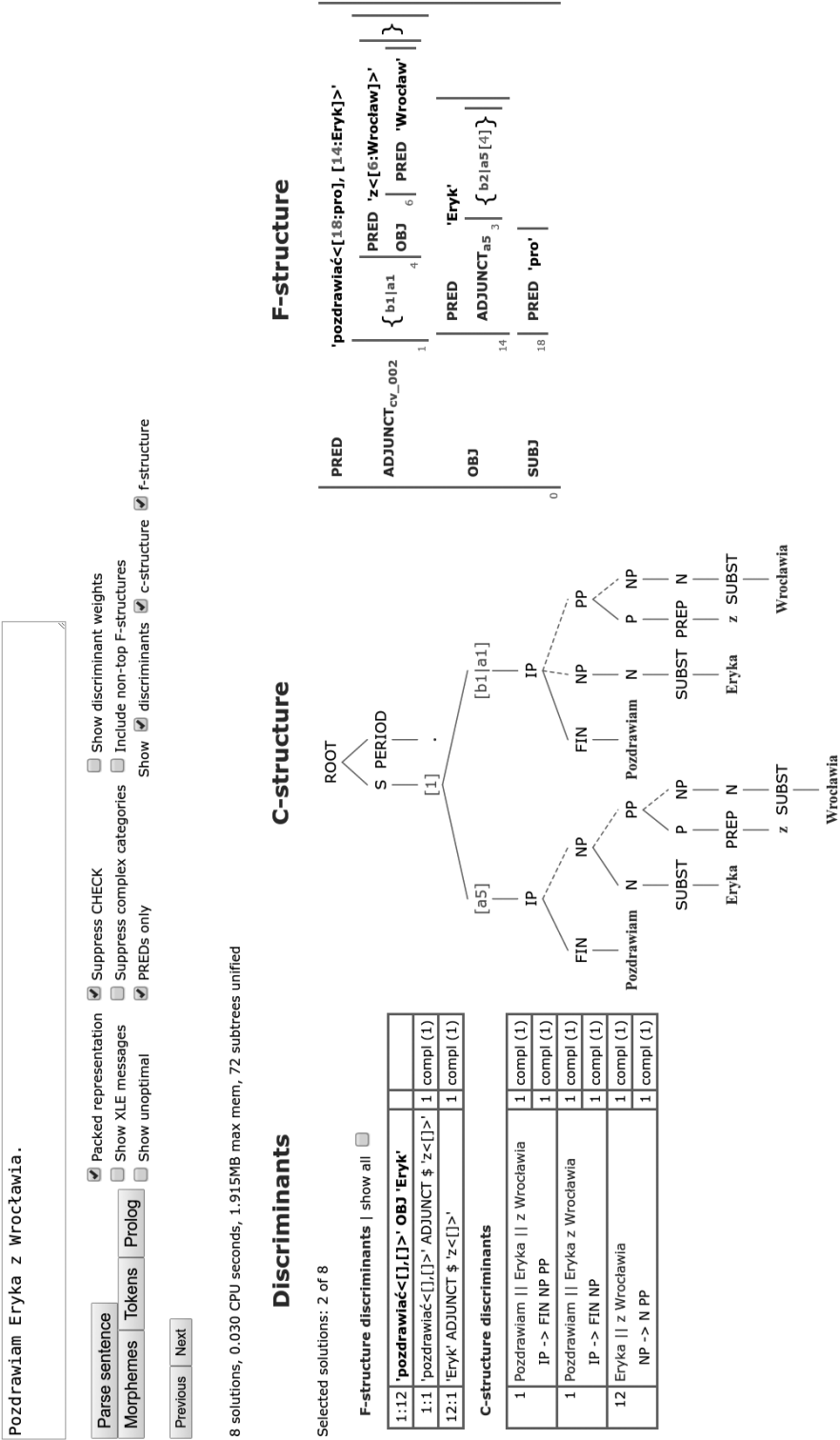
Choć początkowe etapy rozwoju gramatyki POLFIE były zainspirowane wspomnianymi wcześniejszymi gramatykami formalnymi, to obecna postać POLFIE istotnie się od nich różni, zachowując ich cechy pozytywne. Podobnie jak FOJP, ale inaczej niż GFJP, gramatyka POLFIE oparta jest na stabilnej i szeroko znanej teorii lingwistycznej, od dekad rozwijanej przez lingwistów na całym świecie, nieograniczającej się do składni, lecz obejmującej swoim opisem także inne poziomy lingwistyczne, w tym przede wszystkim semantykę. Poszczególne analizy zaimplementowane w gramatyce POLFIE prezentowane są środowisku lingwistycznemu, przede wszystkim – ale nie tylko – od 2012 roku w ramach kolejnych międzynarodowych konferencji LFG. Podobnie natomiast jak GFJP, ale inaczej niż FOJP, gramatyka POLFIE ma szeroki zasięg empiryczny, co jest możliwe między innymi dzięki wykorzystaniu wspomnianych powyżej zewnętrznych zasobów leksykalnych, a także dzięki względnej efektywności platformy implementacyjnej. Analizy gramatyki POLFIE weryfikowane są na ręcznie znakowanym zrównoważonym podzbiorze Narodowego Korpusu Języka Polskiego (NKJP; Przepiórkowski i in. (red.) 2012; <http://nkjp.pl/>), czyli na tzw. korpusie NKJP1M (Degórski, Przepiórkowski 2012; Szałkiewicz, Przepiórkowski 2012); na podstawie gramatyki i NKJP1M powstał – i nadal jest rozwijany – korpus języka polskiego zawierający ręcznie ujednoliconne głębokie rozkłady składniowe (Patejuk, Przepiórkowski 2015b), liczący obecnie około 11 000 zdań¹⁰.

Zarówno sama gramatyka, jak i powstały na jej podstawie bank struktur składniowych są publicznie dostępne (na swobodnej licencji GNU GPLv3). Wersje źródłowe gramatyki znaleźć można na poświęconej jej stronie <http://zil.ipipan.waw.pl/LFG/>, najwygodniej jednak jest testować gramatykę za pomocą interfejsu sieciowego pod adresem <http://iness.mozart.ipipan.waw.pl/iness/xle-web>, opartego na opracowanej na Uniwersytecie w Bergen platformie INESS służącej do – między innymi – wizualizacji wyników gramatyk XLE i do ręcznego ujednolicania analiz (Rosén i in. 2012; <http://iness.uib.no/>).

Platforma ta, dzięki niezwyklej łatwości znajdowania poprawnych analiz wśród tych oferowanych przez gramatykę, została także wykorzystana do budowy wspomnianego banku struktur, dostępnego bezpośrednio ze strony <http://iness.uib.no/11>. W wypadku zdania o więcej niż

10 Dla porównania: oparty na GFJP bank drzew Składnica (<http://zil.ipipan.waw.pl/Składnica/>) liczy obecnie prawie 10 700 zdań, nie zawiera jednak informacji o funkcjach gramatycznych itp.; istotnie większy jest natomiast wspomniany podkorpus NKJP1M, w którym znakowanych składniowo jest ponad 70 tys. zdań, zawiera on jednak tylko płytkie informacje składniowe, bez pełnych rozbiórów zdań.

11 Po wejściu na stronę należy kliknąć na „Treebank Selection”, następnie wybrać „CLARIN-PL”.



jednym rozkładzie składniowym INESS prezentuje wszystkie c-struktury w postaci jednego grafu (jak na środku rys. 1) i wszystkie f-struktury w podobnie upakowanej postaci (jak w prawej części rys. 1). Jednocześnie jednak system analizuje różnice między rozbiorami i prezentuje je w znacznie czytelniejszej postaci tzw. wyróżników (jak w lewej części rys. 1). W zdaniu z rys. 1, *Pozdrawiam Eryka z Wrocławia*, podstawowa niejednoznaczność polega na tym, że fraza *z Wrocławia* może być modyfikatorem czasownika (*pozdrawiam z Wrocławia*) lub rzeczownika (*Eryk z Wrocławia*). Faktycznie liczba rozbiorów jest większa – jest ich osiem – co wynika z tego, że predykat POZDRAWIAĆ może mieć jeden, dwa lub trzy argumenty, a forma rzeczownika ERYK może być analizowana jako dopełnienie bliższe (OBJ) lub jako biernikowy modyfikator czasownika (przez analogię do formy rzeczownika GODZINA w zdaniu *Godzinę pozdrawiałem (kogoś) z Wrocławia*). Rys. 1 przedstawia sytuację, gdy zaakceptowany już został (pogrubiony na rzucie ekranu) wyróżnik 'pozdrawiać<[],[]>' OBJ 'Eryk' mówiący, że predykat POZDRAWIAĆ ma dwa argumenty, a ERYK jest dopełnieniem bliższym tego predykatu. Interpretację, w której fraza *z Wrocławia* modyfikuje czasownik, można następnie wybrać poprzez zaakceptowanie wyróżnika 'pozdrawiać<[],[]>' ADJUNCT \$ 'z<[]>' (według którego fraza *z Wrocławia* jest modyfikatorem, ADJUNCT, predykatu dwuargumentowego POZDRAWIAĆ) lub poprzez odrzucenie wyróżnika 'Eryk' ADJUNCT \$ 'z<[]>' (który mówi, że fraza *z Wrocławia* jest modyfikatorem predykatu ERYK) – obydwie drogi prowadzą do wyboru tej samej f-struktury, w której fraza *z Wrocławia* jest modyfikatorem czasownika.

5. Zakończenie

Gramatyka POLFIE jest nie tylko największą polską gramatyką formalną opartą na ustabilizowanej teorii lingwistycznej, ale także jedną z największych gramatyk rozwijanych w ramach wspomnianego powyżej projektu PARGRAM. W niniejszym artykule naszkicowaliśmy analizy pewnych zjawisk, których inne gramatyki języka polskiego nie obejmują i które – przynajmniej w niektórych wypadkach – byłyby trudne do lingwistycznie adekwatnego opisu w innych podejściach. Gramatyka i powstały na jej podstawie korpus składniowy są nadal rozwijane – obecnie w ramach realizowanego od lipca 2016 roku dwuletniego projektu związanego z europejską inicjatywą CLARIN ERIC.

Bibliografia

- Bresnan J. (red.) 1982: *The mental representation of grammatical relations*, MIT Press Series on Cognitive Theory and Mental Representation, The MIT Press, Cambridge, MA.
- Bresnan J., Asudeh A., Toivonen I., Wechsler S. 2015: *Lexical-functional syntax*, wyd. 2, Blackwell Textbooks in Linguistics, Wiley-Blackwell.
- Butt M., King T.H., Masuichi H., Rohrer C. 2002: *The parallel grammar project*, [w:] N.J. Carroll, R. Sutcliffe (red.), *Proceedings of the workshop on grammar engineering and valuation*, s. 1–7.
- Chomsky N. 1995: *The Minimalist Program*, The MIT Press, Cambridge, MA.
- Dalrymple M. 2001: *Lexical Functional Grammar*, Academic Press, San Diego.
- Degórski Ł., Przepiórkowski A. 2012: *Ręcznie znakowany milionowy podkorpus NKJP*, [w:] A. Przepiórkowski, M. Bańko, R.L. Górski, B. Lewandowska-Tomaszczyk (red.), *Narodowy Korpus Języka Polskiego*, Wydawnictwo Naukowe PWN, Warszawa, s. 51–58.

- Hajnicz E., Patejuk A., Przepiórkowski A., Woliński M. 2016: *Walenty: słownik walencyjny języka polskiego z bogatym komponentem frazeologicznym*, [w:] K. Skwarska, E. Kaczmarek (red.), *Výzkum slovesné valence ve slovanských zemích*, Slovanský ústav AV ČR, Praga, s. 71–102.
- Kallas K. 1993: *Składnia współczesnych polskich konstrukcji współrzędnych*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.
- Landau I. 2013: *Control in generative grammar. A research companion*, Cambridge University Press, Cambridge.
- Patejuk A. 2015: *Unlike coordination in Polish. An LFG account*, rozprawa doktorska, Instytut Języka Polskiego PAN, Kraków.
- Patejuk A. 2016: *Integrating a rich external valence dictionary with an implemented XLE/LFG grammar*, [w:] D. Arnold, M. Butt, B. Crysmann, T.H. King, S. Müller (red.), *The Proceedings of the HeadLex16 Conference*, CSLI Publications, Stanford, CA.
- Patejuk A., Przepiórkowski A. 2012: *Lexico-semantic coordination in Polish*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'12 Conference*, CSLI Publications, Stanford, CA, s. 461–478.
- Patejuk A., Przepiórkowski A. 2014a: *Control into selected conjuncts*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'14 Conference*, CSLI Publications, Stanford, CA, s. 448–460.
- Patejuk A., Przepiórkowski A. 2014b: *In favour of the raising analysis of passivisation*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'14 Conference*, CSLI Publications, Stanford, CA, s. 461–481.
- Patejuk A., Przepiórkowski A. 2014c: *Structural case assignment to objects in Polish*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'14 Conference*, CSLI Publications, Stanford, CA, s. 429–447.
- Patejuk A., Przepiórkowski A. 2015a: *An LFG analysis of the so-called reflexive marker in Polish*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'15 Conference*, CSLI Publications, Stanford, CA, s. 270–288.
- Patejuk A., Przepiórkowski A. 2015b: *Parallel development of linguistic resources. Towards a structure bank of Polish*, „Prace Filologiczne” LXV, s. 255–270.
- Przepiórkowski A., Bańko M., Górski R.L., Lewandowska-Tomaszczyk B. (red.) 2012: *Narodowy Korpus Języka Polskiego*, Wydawnictwo Naukowe PWN, Warszawa.
- Przepiórkowski A., Hajnicz E., Andrzejczuk A., Patejuk A., Woliński M. 2017: *Walenty: gruntowny składniowo-semantyczny słownik walencyjny języka polskiego*, „Język Polski” XCVII, z. 1, s. 30–47.
- Przepiórkowski A., Kupść A., Marciniak M., Mykowiecka A. 2002: *Formalny opis języka polskiego. Teoria i implementacja*, Akademicka Oficyna Wydawnicza EXIT, Warszawa.
- Przepiórkowski A., Patejuk A. 2012a: *On case assignment and the coordination of unlikes. The limits of distributive features*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'12 Conference*, CSLI Publications, Stanford, CA, s. 479–489.
- Przepiórkowski A., Patejuk A. 2012b: *The puzzle of case agreement between numeral phrases and predicative adjectives in Polish*, [w:] M. Butt, T.H. King (red.), *The Proceedings of the LFG'12 Conference*, CSLI Publications, Stanford, CA, s. 490–502.
- Przepiórkowski A., Patejuk A. 2014: *Koordinacja leksykalno-semantyczna w systemie współczesnej polszczyzny (na materiale Narodowego Korpusu Języka Polskiego)*, „Język Polski” XCIV, s. 104–115.
- Przepiórkowski A., Skwarski F., Hajnicz E., Patejuk A., Świdziński M., Woliński M. 2014: *Modelowanie własności składniowych czasowników w nowym słowniku walencyjnym języka polskiego*, „Polonica” XXXIII, s. 159–178.
- Rosén V., De Smedt K., Meurer P., Dyvik H. 2012: *An open infrastructure for advanced treebanking*, [w:] J. Hajič, K. De Smedt, M. Tadić, A. Branco (red.), *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*, Istanbul, s. 22–29.
- Szałkiewicz Ł., Przepiórkowski A. 2012: *Anotacja morfoskładniowa*, [w:] A. Przepiórkowski, M. Bańko, R.L. Górski, B. Lewandowska-Tomaszczyk (red.), *Narodowy Korpus Języka Polskiego*, Wydawnictwo Naukowe PWN, Warszawa, s. 59–96.
- Świdziński M. 1992a: *Gramatyka formalna języka polskiego*, Rozprawy Uniwersytetu Warszawskiego, t. 34, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.
- Świdziński M. 1992b: *Realizacje zdaniowe podmiotu-mianownika, czyli o strukturalnych ograniczeniach selekcyjnych*, [w:] A. Markowski (red.), *Opisać słowa*, Elipsa, Warszawa, s. 188–201.
- Świdziński M. 1993: *Dalsze kłopoty z bezokolicznikiem*, [w:] J. Sambor, J. Linde-Usiekniewicz, R. Huszcza (red.), *Językoznawstwo synchroniczne i diachroniczne*, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa, s. 303–314.
- Witkoś J., Cegłowski P., Snarska A., Żychliński S. 2011a: *Minimalist facets of control. An English-Polish comparative overview of gerunds and infinitives*, Wydawnictwo Naukowe Uniwersytetu im. A. Mickiewicza, Poznań.
- Witkoś J., Cegłowski P., Snarska A., Żychliński S. 2011b: *Wybrane aspekty derywacji i interpretacji wyrażen bezokolicznikowych. Minimalistyczne studium porównawcze polsko-angielskie*, Wydawnictwo Naukowe Uniwersytetu im. A. Mickiewicza, Poznań.

- Woliński M. 2004: *Komputerowa weryfikacja gramatyki Świdzińskiego*, praca doktorska, Instytut Podstaw Informatyki Polskiej Akademii Nauk, Warszawa.
- 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.

Summary

POLFIE: a contemporary formal grammar of Polish

Keywords: grammar, LFG, Polish, syntax.

This paper presents the new formal grammar of Polish, POLFIE, couched in Lexical Functional Grammar. The paper introduces the basics of LFG, discusses the way of representing syntactic structures in LFG and demonstrates the power of the LFG formalism on the basis of the analysis of selected phenomena of Polish in POLFIE. The paper shows that the LFG theory makes it possible to provide a formalised, adequate and linguistically deep description of syntactic phenomena such as: syntactic control, predicative arguments, the haplology of *się*, unbounded dependencies and coordination; it also takes into account interactions between these. POLFIE is accompanied by a computer implementation in XLE.