LingVaria
X (2015) | 2(20)
doi: 10.12797/LV.10.2015.20.11

Adam Przepiórkowski
Instytut Podstaw Informatyki PAN,
Uniwersytet Warszawski, Warszawa
adamp@ipipan.waw.pl

INŻYNIERIA LINGWISTYCZNA A OBECNA SYTUACJA JĘZYKOZNAWSTWA POLSKIEGO

Słowa klucze: lingwistyka informatyczna, lingwistyka formalna, językoznawstwo w Polsce
Keywords: computational linguistics, formal linguistics, linguistics in Poland

Tytuł niniejszego artykułu[1] nawiązuje do opublikowanego w poprzednim numerze „LingVariów" artykułu Piotra Żmigrodzkiego oceniającego stan językoznawstwa polskiego z punktu widzenia twórców *Wielkiego słownika języka polskiego PAN* (WSJP). Podobnie jak tamten artykuł, ten również ma charakter bardziej publicystyczny niż naukowy. Chciałbym w nim zwrócić uwagę na to, że podobna jest ocena obecnej sytuacji językoznawstwa polskiego z punktu widzenia dziedziny na pograniczu lingwistyki i informatyki – dziedziny, którą przyjęło się nazywać *lingwistyką informatyczną* lub *inżynierią lingwistyczną*.

„Leksykografia stanowi […] kanał upowszechniania wyników analiz lingwistycznych ściśle naukowych" (Żmigrodzki 2015: 110), natomiast inżynieria lingwistyczna takie ściśle naukowe analizy wykorzystuje do tworzenia narzędzi przetwarzania tekstów (na przykład tzw. parserów, czyli analizatorów składniowych) i coraz

1 Za uwagi do wcześniejszej wersji artykułu dziękuję serdecznie Januszowi S. Bieniowi, Magdalenie Danielewiczowej, Włodzimierzowi Gruszczyńskiemu, Agnieszce Patejuk i Markowi Świdzińskiemu. Podziękowania te nie implikują oczywiście bezwarunkowego poparcia wyżej wymienionych dla wszystkich stawianych tu tez.

liczniejszych aplikacji (systemów tłumaczenia automatycznego czy też bardziej inteligentnych wyszukiwarek internetowych). W kierowanym przeze mnie Zespole Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN jest i było realizowanych wiele projektów krajowych i europejskich[2], związanych z najróżniejszymi poziomami wiedzy lingwistycznej: od sygnałów akustycznych mowy, poprzez system morfoskładniowy i składniowy polszczyzny, do semantyki leksykalnej i kompozycjonalnej oraz zjawisk takich jak koreferencja i metafora; prowadzimy także intensywne prace leksykograficzne, o których więcej poniżej. W obecną sytuację językoznawstwa polskiego mam także wgląd z zupełnie innej perspektywy, jako członek Zarządu Polskiego Towarzystwa Językoznawczego w drugiej już kadencji, uczestniczący w procesie oceny i akceptacji zgłoszeń nadsyłanych na zjazdy PTJ. Z obu tych perspektyw wyłania się – podobnie jak z perspektywy leksykografii współczesnej – „obraz dość pesymistyczny" (Żmigrodzki 2015: 113).

Zacznę od obszaru mi najbliższego, a więc od składni. Jakie podejścia do składni dominują na świecie? Żeby nie być gołosłownym, posłużę się czterema kompendiami składniowymi z ostatnich lat. Najstarsze z nich, z roku 2011, to *Non-Transformational Syntax: Formal and Explicit Models of Grammar* (Borsley, Börjars 2011). Główne teorie składniowe, które zostały tam uznane za godne opisania, to Head-driven Phrase Structure Grammar (HPSG: Pollard, Sag 1994), Lexical Functional Grammar (LFG: Bresnan 1982; Dalrymple 2001) i gramatyki kategorialne (Ajdukiewicz 1935; Steedman 1996) – każdej z nich poświęcono dwa rozdziały. Parę innych rozdziałów dotyczy ogólnie podejść opartych na ograniczeniach (ang. *constraint-based*), takich jak wspomniane teorie HPSG i LFG, a jeden – autorstwa Raya Jackendoffa – poświęcony jest alternatywnym podejściom do Programu Minimalistycznego (Chomsky 1995). Z roku 2014 pochodzi pozycja *Theories of Syntax: Concepts and Case Studies* (Kuiper, Nokes 2014). Teorie tu omówione to Systemic Functional Grammar (Halliday 1994), znowu LFG i znowu podejścia generatywno-transformacyjne Chomsky'ego. W kompendium *Syntax – Theory and Analysis. An International Handbook* (Kiss, Alexiadou 2015) w części *Syntactic Models* omówione zostały te same teorie co w wydawnictwie z roku 2011, a także Construction Grammar (Goldberg 1995), gramatyki zależnościowe oraz modelowanie składni w Optimality Theory (Prince, Smolensky 1993). W końcu w mającym dopiero się ukazać w tym roku, ale już dostępnym w postaci PDF[3], podręczniku *Grammatical Theory: From Transformational Grammar to Constraint-Based Approaches* (Müller 2015a) znowu omówione zostały teorie Chomsky'ego, HPSG, LFG, gramatyki kategorialne, Construction Grammar i gramatyki zależnościowe, a także Tree Adjoining Grammar (Joshi 1985) i podejście będące prekursorem HPSG, czyli Generalized Phrase Structure Grammar (GPSG; Gazdar et al. 1985). Daje to następujący ranking:

2 Listę i opisy tych projektów zob. http://zil.ipipan.waw.pl.

3 Zob. http://langsci-press.org/catalog/book/25.

gramatyki generatywno-transformacyjne	4
Lexical Functional Grammar	4
Head-driven Phrase Structure Grammar	3
gramatyki kategorialne	3
gramatyki zależnościowe	2
Generalized Phrase Structure Grammar	1
Tree Adjoining Grammar	1
Optimality Theory	1
Systemic Functional Grammar	1
Construction Grammar	1

Porównajmy te dane z obrazem syntaktyki polskiej w artykule Macieja Grochowskiego (2012). Poszczególne punkty artykułu omawiają: składnię tradycyjną (przede wszystkim Klemensiewiczowską), składnię strukturalną w wersji klasycznej (spod znaku Kuryłowicza), tzw. składnię semantyczną (Karolak, Bogusławski) i składnię formalną, do której zaliczane są „składnia dystrybucyjna" (Misz, Szupryczyńska, Kallas, Saloni, Świdziński i in.) oraz składnia „ukształtowana pod wpływem generatywizmu" (Polański, Bobrowski, ale także Wojtasiewicz i Bellert)[4]. Jak łatwo zauważyć, w przecięciu tych dwóch podsumowań – naszego na podstawie czterech współczesnych kompendiów składni i Grochowskiego, opisującego sytuację w Polsce – znajdują się wyłącznie gramatyki generatywno-transformacyjne, i to tylko przy zbiorowym potraktowaniu wszelkich odmian generatywno-transformacyjnych: Grochowski wspomina wyłącznie o pracach odnoszących się do wczesnych modeli Chomsky'ego, aktualne kompendia zaś omawiają modele z lat 90. i późniejszych (Program Minimalistyczny – we wszystkich czterech wydawnictwach) oraz z lat 80. (Teoria Rządu i Wiązania – w dwóch wypadkach).

Oczywiście fakt, że myśl składniowa rozwija się w Polsce w dużej mierze w oderwaniu od tendencji światowych, nie oznacza automatycznie, że polskie językoznawstwo stoi na niższym poziomie. Oznacza jednak, że syntaktyka w Polsce nie podąża za ogólnym trendem polegającym na znacznie większym stopniu formalizacji analiz: zwróćmy uwagę na to, że wszystkie najczęściej wymieniane w kompendiach teorie to teorie mocno sformalizowane: tak jest niewątpliwie w wypadku HPSG i gramatyk kategorialnych, ale także w wypadku LFG i – obecnie w znacznie mniejszym stopniu – gramatyk generatywno-transformacyjnych. W odróżnieniu od przytłaczającej większości analiz polonistycznych analizy sformułowane w ramach tych teorii mogą

4 Pomijam tu krótkie punkty poświęcone „składni poziomu meta" i składni w językoznawstwie kognitywnym.

być mniej lub bardziej bezpośrednio przeniesione do komputerowej implementacji gramatyki danego języka[5].

Jak pisze Żmigrodzki:

> Badaniem polszczyzny zajmują się u nas w tej chwili zasadniczo trzy kolektywy badawcze: językoznawcy poloniści, językoznawcy neofilologiczni (przede wszystkim chyba angliści) oraz przedstawiciele nauk technicznych, rozwijający tzw. lingwistykę komputerową czy inżynierię językową (Żmigrodzki 2015: 115).

Przynajmniej w zakresie sformalizowanych prac składniowych poloniści oddają pole pozostałym dwóm grupom. Nie licząc unikatowej we współczesnym generatywizmie transformacyjnym spójnej analizy dużego fragmentu danego języka naturalnego dokonanej przez polonistę, Ireneusza Bobrowskiego (2005)[6], intensywne – choć zwykle, z powodów zupełnie oczywistych, kontrastowe raczej niż czysto polonistyczne – prace w ramach tego paradygmatu prowadzone są głównie na anglistykach: poznańskiej, wrocławskiej, lubelskiej, krakowskiej i innych. Także jedyna znana mi monografia dotycząca analizy pewnych zjawisk języka polskiego w teorii Construction Grammar powstała na anglistyce (warszawskiej; Szymańska 2000), a bodajże jedyna dotychczas obroniona w Polsce praca doktorska w ramach teorii LFG została napisana na… orientalistyce (znowu warszawskiej; Olejarnik 2009).

Obraz, jaki zastaje lingwista informatyczny próbujący opracować system analizy składniowej – albo składniowo-semantycznej – polszczyzny, jest więc następujący: z jednej strony mamy do dyspozycji niewątpliwe osiągnięcia szkoły składni dystrybucyjnej, ale osiągnięcia już w dużej mierze wyeksploatowane w formalnych powierzchniowo-składniowych gramatykach Szpakowicza (1978, 1983) i Świdzińskiego (1992)[7], a także we wczesnych wersjach rozwijanej obecnie polskiej gramatyki LFG (więcej o tym poniżej), z drugiej zaś strony – mniej lub bardziej formalne analizy wybranych zjawisk języka polskiego w ramach teorii generatywno-transformacyjnych, które nie poddają się łatwo implementacji. Aby więc móc skorzystać ze współczesnych teorii składniowych dających silniejsze i lepiej sformalizowane narzędzia analizy, takich jak HPSG czy LFG, lingwista informatyczny musi przejąć na siebie obowiązki lingwisty polonistycznego i sam zacząć tworzyć opisy w ich ramach.

---

5 Stwierdzenie to nie dotyczy gramatyk generatywno-transformacyjnych, gdyż wykorzystywane przez nie pojęcie transformacji nie jest łatwe do efektywnego zaimplementowania, a także z powodu dużego rozdrobnienia analiz i częstego ignorowania zjawisk wykraczających poza coraz węziej rozumiany *trzon* (ang. *core*) języka (zob. Müller 2015b).

6 Choć monografia ukazała się w roku 2005, a jej wcześniejsza dwutomowa wersja w latach 90. XX w., to wykorzystywany jest w niej formalizm generatywno-transformacyjny z lat 60., czyli tzw. model standardowy Chomsky'ego (z elementami GPSG).

7 Implementacja tej drugiej – nadal rozwijana – opisana jest w obronionej w IPI PAN pracy doktorskiej Wolińskiego (2004).

W polskiej lingwistyce informatycznej taka sytuacja miała miejsce co najmniej trzykrotnie. Na przełomie lat 90. XX w. i początku wieku XXI stworzony został formalny – i spójny wewnętrznie – opis wybranych zjawisk języka polskiego (w tym niektórych nieanalizowanych wcześniej tak szczegółowo, jak na przykład nadawanie przypadka strukturalnego i wiązanie zaimków anaforycznych) w ramach teorii Head-driven Phrase Structure Grammar. Opis ten – czysto teoretyczny, choć mocno sformalizowany i poparty próbą implementacji komputerowej – został opublikowany w postaci monografii (Przepiórkowski et al. 2002), a towarzyszyły mu cztery doktoraty, także o charakterze formalno-lingwistycznym raczej niż implementacyjnym, oraz liczne artykuły, również opublikowane w materiałach konferencji slawistycznych. W środowisku polonistycznym prace te nie tylko pozostały praktycznie bez odzewu[8], co zapewne nie dziwi, gdy weźmie się pod uwagę stopień ich formalnego zaawansowania, ale też – jeśli już zostają zauważone – funkcjonują w świadomości polonistów jako wyłącznie inżynieryjne, a nie teoretycznolingwistyczne[9]. W podobnym czasie został napisany na Politechnice Poznańskiej i obroniony w IPI PAN doktorat proponujący zależnościowy opis szerokich fragmentów języka polskiego (Obrębski 2002) – o ile mi wiadomo, także on pozostał praktycznie niezauważony. Obecnie trwają natomiast w IPI PAN prace zmierzające do stworzenia – w ramach teorii Lexical Functional Grammar – gramatyki składniowo-semantycznej o szerokim pokryciu empirycznym (Patejuk, Przepiórkowski 2012). W wypadku tych prac wkładamy więcej świadomego wysiłku w dotarcie do polonistów (np. Patejuk, Przepiórkowski w druku), może więc tym razem uda się nawiązać jakiś dialog, choć nie nastraja optymistycznie obserwacja, że

> […] widoczna jest […] nieufność językoznawców do lingwistów „technicznych", a nawet ostentacyjne lekceważenie tej gałęzi badań nad językiem, tym bardziej przykre, że wyrażane przez osoby, które z racji pełnionych funkcji raczej powinny dążyć do zjednoczenia sił wszystkich badaczy języka (Żmigrodzki 2015: 115).

Podobnie wygląda sytuacja w wypadku semantyki, w której powstające opisy są zwykle zbyt mało sformalizowane czy precyzyjne, by mogły być bezpośrednio wykorzystane w przetwarzaniu języka. Jak wielu polskich semantyków potrafi się posługiwać rachunkiem lambda? Jak wielu zna podstawy logiki intensjonalnej wy-

8 Aby nie być gołosłownym, podaję przykład: Kiklewicz (2012) omawia zjawisko tzw. haplologii *się*, nie odwołując się do bodajże pierwszego – i chyba nadal jedynego – wnikliwego opisu tego zjawiska (Kupść 1999, 2000), sformułowanego w terminach teorii HPSG. Wspomniane tu i poniżej prace HPSG i LFG spotkały się natomiast z odzewem poza Polską, o czym mogą świadczyć publicznie dostępne dane z Google Scholar.

9 Na przykład: „Koncepcje składni formalnej są wykorzystywane przez zespoły […] zajmujące się komputerowym generowaniem języka […] (Przepiórkowski, Kupść, Marciniak, Mykowiecka 2002)" (Grochowski 2012: 149). Cytat ten jest mylący także dlatego, że w tym czasie nie zajmowaliśmy się generowaniem języka, a jedynie jego analizą.

korzystanej przez Richarda Montague'a – ucznia Alfreda Tarskiego – do pokazania, że opis semantyki kompozycjonalnej podlega całkowitej formalizacji? Jak wielu potrafi wykorzystać w swoich analizach pojęcie kwantyfikatora uogólnionego, wprowadzone zresztą przez polskiego matematyka i logika Andrzeja Mostowskiego? Oczywiście nie chodzi o to, by wszystkie prace formalizować na każdym ich etapie, martwi natomiast prawie całkowity brak takich formalizacji. Badanie języków z wykorzystaniem tych – powszechnie stosowanych w semantyce światowej – pojęć jest raczej domeną wydziałów matematyczno-informatycznych i filozoficznych niż filologicznych.

Konieczność tworzenia odpowiednio sformalizowanych opisów składni czy semantyki to niejedyny obszar, w którym lingwiści informatyczni wyręczają lingwistów polonistycznych i leksykografów. Na Politechnice Wrocławskiej tworzony jest – zgodnie z metodologią słownika angielskiego Princeton WordNet – wielki słownik semantyczny języka polskiego *Słowosieć* (Piasecki, Szpakowicz, Broda 2009). Słownik ten grupuje jednostki leksykalne w zbiory synonimów (tzw. *synsety*) i określa relacje semantyczne pomiędzy tymi zbiorami – hiperonimię, meronimię i wiele innych. Takie wordnety istnieją już dla wielu języków; polski zaczął być tworzony dosyć późno, ale jest obecnie jednym z największych i najbardziej zaawansowanych tego typu zasobów na świecie. Innym przykładem nowoczesnego słownika tworzonego poza naturalnym środowiskiem leksykograficznym jest *Walenty* (zob. Hajnicz et al. w druku i inne prace tam cytowane), polski słownik walencyjny o znacznie większym pokryciu empirycznym i bardziej zaawansowanym aparacie pojęciowym niż największy do niedawna słownik tego typu, tj. *Słownik syntaktyczno-generatywny czasowników polskich* (SSGC: Polański 1980–1992). Oczywiście w obu wypadkach prace lingwistyczne wykonywane są przede wszystkim przez odpowiednio przeszkolonych lingwistów, a nie przez informatyków[10]. Informatycy zapewniają natomiast narzędzia ułatwiające i przyspieszające takie prace, a także dbają o zachowanie odpowiedniego stopnia precyzji i spójności.

Podsumowując ten wątek, stwierdzam, że lingwistyka polska, rozwijana przez dekady w oderwaniu od osiągnięć lingwistyki światowej, w stopniu dalece niewystarczającym dostarcza materiałów – czy to analiz, czy zasobów – potrzebnych do stworzenia bardziej sformalizowanych i implementowalnych opisów polszczyzny. Gdy zaś już takie zasoby powstają, często pozostają zamknięte w szufladach lub – w najlepszym razie – w opasłych tomach, zamiast być udostępniane w postaci elek-

10 Podział na „lingwistów" i „informatyków" należy zresztą traktować jako bardzo nieostry, co najlepiej widać po składzie osobowym Zespołu Inżynierii Lingwistycznej IPI PAN: ja sam mam magisterium i habilitację informatyczne, ale doktorat z lingwistyki teoretycznej (i czuję się bardziej lingwistą niż informatykiem), niektórzy członkowie zespołu ukończyli studia informatyczne (i nierzadko wykonują pracę lingwistyczną), a inni – lingwistyczne lub pokrewne (i czasami wykonują prace informatyczne). Jedna osoba ukończyła w Niemczech studia z „lingwistyki informatycznej" – jest bardziej lingwistą czy informatykiem?

tronicznej, znacznie łatwiejszej do wykorzystania nie tylko w przetwarzaniu języka, ale także w codziennej praktyce lingwistycznej. Znakomitym przykładem jest tu wspomniany już SSGC, który do postaci elektronicznej został przetworzony przez i z inicjatywy (co dziwi) informatyków[11].

Powyższe kilka akapitów może brzmieć jak narzekania sfrustrowanego lingwisty informatycznego. Tak nie jest – inżynieria lingwistyczna rozwija się w Polsce wyjątkowo prężnie, przede wszystkim w ośrodkach warszawskim, wrocławskim i poznańskim, ale także w innych miejscach, a braki wynikające ze stanu lingwistyki polskiej udaje nam się uzupełnić w ramach własnych prac. Naszkicowana powyżej sytuacja stanowi raczej zagrożenie dla samej lingwistyki uprawianej w Polsce, już obecnie zmarginalizowanej na świecie. Znowu trafnie diagnozuje to Żmigrodzki (2015: 115), pisząc, że „[…] nie jest również możliwy rozwój współczesnych badań nad polszczyzną bez wsparcia ze strony informatyków, zwłaszcza zaś informatyków o nachyleniu lingwistycznym"; coraz częściej zauważają to także inni językoznawcy, jak wynika choćby z odpowiedzi na ankietę jubileuszową „Języka Polskiego" (JP 2013) i z artykułu Renaty Przybylskiej (2013: 29): „[…] bodajże największa przyszłość stoi przed tym działem językoznawstwa, który łączy się z technologią informatyczną". Z satysfakcją obserwuję coraz większe wykorzystanie Narodowego Korpusu Języka Polskiego (NKJP; Przepiórkowski et al. 2012) nie tylko w pracach leksykograficznych (przede wszystkim w procesie tworzenia WSJP), ale także w badaniach składniowych, semantycznych i innych; w pełni zgadzam się także ze Żmigrodzkim (2015: 114), że: „Kontynuacja prac nad NKJP […] jest zadaniem o kluczowym znaczeniu z punktu widzenia przyszłości […] całego polskiego językoznawstwa XXI w." – prace takie powinny mieć w miarę możliwości stałe finansowanie. Wpływ metod przetwarzania języka na rozwój językoznawstwa będzie jednak coraz znaczniej wykraczał poza takie trywialne (z punktu widzenia lingwistyki informatycznej) zastosowania korpusowe, a coraz częściej inżynieria lingwistyczna będzie dostarczała metod automatycznego odkrywania wiedzy lingwistycznej danej w tekstach tylko *implicite*. Jako prosty przykład niech posłuży zadanie wykrywania kolokacji w tekstach – już teraz żaden szanujący się lingwista-leksykograf, mając za zadanie opracowanie słownika kolokacji, powiedzmy, ekonomicznych, nie analizowałby tekstów dziedzinowych ręcznie, tylko skorzystałby z jednego z wielu dostępnych programów znajdujących takie kolokacje automatycznie. Oczywiście wynik działania takich programów należy uporządkować i zinterpretować – i tu lingwista jest (przynajmniej obecnie) niezbędny – ale większość pracy wykonuje program zawierający odpowiednie

11 Zob. http://zil.ipipan.waw.pl/SGDPV. Historia powstania tej wersji jest dosyć długa, ale w największym skrócie: słownik został przepisany do bazy MS Access w roku 2000 w ramach prac na Politechnice Śląskiej, nie mógł jednak być w tej postaci rozpowszechniany z powodu braku odpowiednich licencji. Taką zgodę na rozpowszechnianie wersji elektronicznej udało się uzyskać IPI PAN; sam słownik został wyczyszczony i przetworzony do postaci XML w ramach projektu europejskiego (sic!) CESAR realizowanego w ZIL IPI PAN.

wzory statystyczne i umiejący przypisać charakterystykę morfosyntaktyczną słowom tekstowym. Już ten prosty przykład pokazuje, że współczesne metody pozwalają znajdować informację tam, gdzie nie była ona podana *explicite*: nie oczekujemy, że w tekstach wejściowych będą zaznaczone kolokacje – zamiast tego program działa na tekstach nieoznakowanych i sam takie kolokacje znajduje. Bardziej zaawansowanym przykładem są systemy wykrywające zależności walencyjne w tekstach, które są oznakowane jedynie morfoskładniowo, a nie składniowo. Takie programy komputerowe powstają od około połowy lat 90., a w kontekście języka polskiego systemy takie zostały opracowane w ramach projektu[12] kierowanego przeze mnie w latach 2005–2008. Kolejnym dobrym przykładem może być zastosowanie metod automatycznego znajdowania wyrazów semantycznie zbliżonych do danego – choć metody takie nadal zbyt często mylą się w określaniu, czy dane dwa wyrazy pozostają w relacji hiperonimii, synonimii czy może antonimii, by możliwa była pełna automatyzacja procesu konstrukcji słownika semantycznego typu wordnet, to na tyle dokładnie znajdują one wyrazy znaczeniowo spokrewnione, by możliwe było istotne przyspieszenie pracy leksykografa i zwiększenie jakości wyników jego pracy. Inżynieria lingwistyczna w Polsce osiągnęła więc stan, w którym coraz mniej jest zależna od lingwistyki teoretycznej czy polonistycznej, coraz więcej może natomiast takiej „czystej lingwistyce" zaoferować.

Wydaje się więc, że językoznawstwo polskie stanęło obecnie na rozdrożu. Może ono kontynuować obserwowany obecnie w wielu – choć nie we wszystkich! – obszarach chów wsobny, gdzie poszczególne towarzystwa wzajemnej adoracji rozwijają tę lub tamtą teorię bez oglądania się nie tylko na dokonania lingwistyki światowej i informatycznej, lecz także na inne towarzystwa wzajemnej adoracji. Ale być może jest to dobry moment, żeby – jak pisze Żmigrodzki (2015: 115) – „zastanowić się nad rolą tzw. środowiska językoznawczego w rozwoju naszej dyscypliny teraz i w przyszłości". Niestety, nie napawa optymizmem zarysowany przez Żmigrodzkiego (ibid.: 116) stan kształcenia młodych językoznawców zakończony konkluzją: „obecny absolwent polonistyki […] na pewno nie jest przygotowany do pracy leksykograficznej i naukowej".

Mimo tej diagnozy chciałbym jednak zamknąć niniejszy artykuł nutą optymizmu. W IPI PAN od pierwszych lat obecnego wieku intensywnie współpracujemy z lingwistami – przeważnie polonistami – i jest to zwykle współpraca satysfakcjonująca dla obu stron, nawet jeżeli dotyczy tak żmudnego zadania jak ręczne ujednoznacznianie informacji morfosyntaktycznych w korpusach (najpierw w Korpusie IPI PAN, potem w NKJP). Wielokrotnie spotykaliśmy się z uwagami, także ze strony doktorantów studiów polonistycznych, że dopiero intensywna praca nad autentycznymi tekstami, w których każde słowo musi być zinterpretowane w sposób precyzyjny, uświadomiła im z jednej strony bogactwo polszczyzny, a z drugiej

12 Zob. http://nlp.ipipan.waw.pl/PPJP/.

strony potrzebę sformalizowanego podejścia do języka, mimo pewnej arbitralności takich formalizacji w niektórych przypadkach. Owocną współpracę prowadziliśmy i prowadzimy nie tylko z polonistyką warszawską, lecz także z doktorantami i pracownikami Katedry Lingwistyki Formalnej, z różnymi zakładami Instytutu Języka Polskiego PAN, z frazeologami z Uniwersytetu Warmińsko-Mazurskiego w Olsztynie oraz z lingwistami z wielu innych ośrodków. Na niedawno powstałych na Uniwersytecie Warszawskim Studiach Kognitywistycznych[13] już na poziomie licencjackim nauczana jest lingwistyka strukturalistyczna spod znaku de Saussure'a, ale także współczesna składnia formalna (LFG, z elementami składni dystrybucyjnej i zależnościowej) i współczesna semantyka formalna; absolwenci tych studiów będą zapewne lepiej przygotowani do współpracy z lingwistami informatycznymi – i do nowoczesnych badań lingwistycznych w ogóle – niż absolwenci niejednej polonistyki czy filologii. Zaryzykowałbym więc stwierdzenie, że powstaje pewien ekosystem, w którym nawiązany został dialog pomiędzy językoznawstwem informatycznym a językoznawstwem teoretycznym i polonistycznym, i to ekosystem znacznie bardziej zróżnicowany niż modelowa, ale dotycząca jedynie morfoskładni i składni powierzchniowej współpraca informatyków (Bień, Szpakowicz) i lingwistów (Saloni, Świdziński, Gruszczyński) na Uniwersytecie Warszawskim w latach 70. Pozostaje mieć nadzieję, że taka bliższa integracja środowiska możliwa jest także na poziomie ogólnokrajowym.

Literatura

Ajdukiewicz K., 1935, *Die syntaktische Konnexität*, „Studia Philosophica" 1, s. 1–27.

Bobrowski I., 2005, *Składniowy model polszczyzny*, Kraków.

Borsley R.D., Börjars K. (red.), 2011, *Non-Transformational Syntax. Formal and Explicit Models of Grammar*, Oxford.

Bresnan J. (red.), 1982, *The Mental Representation of Grammatical Relations*, Cambridge, MA.

Chomsky N., 1995, *The Minimalist Program*, Cambridge, MA.

Dalrymple M., 2001, *Lexical Functional Grammar*, San Diego, CA.

Gazdar G. et al., 1985, *Generalized Phrase Structure Grammar*, Cambridge, MA.

Goldberg A.E., 1995, *Constructions. A Construction Grammar Approach to Argument Structure*, Chicago, IL.

Grochowski M., 2012, *Główne kierunki badań syntaktycznych w Polsce w drugiej połowie XX wieku i na początku XXI wieku*, [w:] M. Grochowski (red.), *Językoznawstwo w Polsce. Kierunki badań i perspektywy rozwoju*, Warszawa, s. 139–155.

Hajnicz E. et al., w druku, *Internetowy słownik walencyjny języka polskiego oparty na danych korpusowych*, „Prace Filologiczne" LXV.

Halliday M.A., 1994, *An Introduction to Functional Grammar*, Londyn.

13 Zob. http://kognitywistyka.uw.edu.pl/.

„Język Polski" XCIII, z. 1, 2013.
Joshi A.K., 1985, *Tree adjoining grammars. How much context sensitivity is required to provide reasonable structural descriptions?*, [w:] D. Dowty, L. Karttunen, A.M. Zwicky (red.), *Natural Language Parsing. Psychological, Computational, and Theoretical Perspectives*, Cambridge.
Kiklewicz A., 2012, *Sługa dwóch panów. Wyrazek* się *w aspekcie funkcjonalnym*, „Prace Językoznawcze" XIV, s. 135–146.
Kiss T., Alexiadou A. (red.), 2015, *Syntax – Theory and Analysis. An International Handbook*, t. 1–3, Berlin.
Kuiper K., Nokes J., 2014, *Theories of Syntax. Concepts and Case Studies*, Basingstoke.
Kupść A., 1999, *Haplology of the Polish reflexive marker*, [w:] R.D. Borsley, A. Przepiórkowski (red.), *Slavic in Head-Driven Phrase Structure Grammar*, Stanford, CA, s. 91–124.
Kupść A., 2000, *An HPSG Grammar of Polish Clitics*, Warszawa–Paryż, rozprawa doktorska, mps.
Müller S., 2015a, *Grammatical Theory. From Transformational Grammar to Constraint-Based Approaches*, Berlin.
Müller S., 2015b, *The CoreGram project. Theoretical linguistics, theory development and verification*, „Journal of Language Modelling" 3(1), s. 21–86.
Obrębski T., 2002, *Automatyczna analiza składniowa języka polskiego z wykorzystaniem gramatyki zależnościowej*, Warszawa, rozprawa doktorska, mps.
Olejarnik M., 2009, *Complex Predicates in Swahili. An LFG Approach*, Warszawa, rozprawa doktorska, mps.
Patejuk A., Przepiórkowski A., 2012, *Towards an LFG parser for Polish. An exercise in parasitic grammar development*, [w:] *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, Istanbul, s. 3849–3852.
Patejuk A., Przepiórkowski A., w druku, *Parallel development of linguistic resources. Towards a structure bank of Polish*, „Prace Filologiczne" LXV.
Piasecki M., Szpakowicz S., Broda B., 2009, *A Wordnet from the Ground Up*, Wrocław.
Polański K. (red.), 1980–1992, *Słownik syntaktyczno-generatywny czasowników polskich*, Wrocław–Kraków.
Pollard C., Sag I.A., 1994, *Head-driven Phrase Structure Grammar*, Chicago, IL.
Prince A., Smolensky P., 1993, *Optimality Theory. Constraint interaction in generative grammar*, RuCCS Technical Report 2, Picateway, NJ.
Przepiórkowski A. et al., 2002, *Formalny opis języka polskiego. Teoria i implementacja*, Warszawa.
Przepiórkowski A. et al. (red.), 2012, *Narodowy Korpus Języka Polskiego*, Warszawa.
Przybylska R., 2013, *Językoznawstwo praktyczne czy stosowane – jaka przyszłość*, „Polonica" XXXIII, s. 25–31.
Steedman M., 1996, *Surface Structure and Interpretation*, Cambridge, MA.
Szpakowicz S., 1978, *Automatyczna analiza składniowa zdań pisanych*, Warszawa, rozprawa doktorska, mps.
Szpakowicz S., 1983, *Formalny opis składniowy zdań polskich*, Warszawa.
Szymańska I., 2000, *A Construction Grammar Account of the Reflexive* się *in Polish*, Warszawa, rozprawa doktorska, mps.
Świdziński M., 1992, *Gramatyka formalna języka polskiego*, Warszawa.

Woliński M., 2004, *Komputerowa weryfikacja gramatyki Świdzińskiego*, Warszawa, rozprawa doktorska, mps.
Żmigrodzki P., 2015, *Wielki słownik języka polskiego PAN a obecna sytuacja językoznawstwa polskiego*, „LingVaria" X nr 1(19), s. 109–119.

Linguistic engineering and the current situation of Polish linguistics
Summary

The thesis of this reply to Piotr Żmigrodzki's pessimistic diagnosis (published in the previous issue of „LingVaria") of Polish linguistics from the perspective of modern lexicography is that the diagnosis from the perspective of linguistic engineering must be equally pessimistic – in fact, even more so. We argue that syntax and – to some extent – semantics are developed in Poland in isolation from developments outside Poland and the resulting analyses too often do not meet the usual criteria of preciseness and formal rigour. We end with an optimistic note showing that fruitful cooperation between computational and theoretical linguists is possible – even if rare – in Poland.