

Od informatyki i jej zastosowań do światopoglądu informatycznego,
praca zbiorowa pod red. Pawła Stacewicza, Oficyna Wydawnicza Politechniki
Warszawskiej, Warszawa 2015

Rozdział 6

Informatyka szansą na rozwój naukowej lingwistyki¹

André Włodarczyk

Université Charles de Gaulle (Lille 3),
Paris-Sorbonne (Paris 4)

Streszczenie. Lingwistykę drugiej połowy XX wieku cechuje dążenie do w miarę ścisłego opisu zjawisk językowych. Wprawdzie już pod koniec XIX stulecia F. de Saussure poszukiwał takiej definicji języka, dzięki której językoznawstwo mogłoby się stać dyscypliną autonomiczną. Jednakże dopiero pod wpływem matematyki i informatyki zrodziły się tendencje, by opisywać zjawiska językowe przy pomocy narzędzi formalnych.

Początek XXI wieku zdaje się wskazywać na to, że zarówno lingwistyka strukturalna jak i informatyczna korzysta ze zdobyczy informatyki, która umożliwia im przebadanie coraz bogatszych baz danych materiałowych dotyczących poszczególnych języków. Wobec wielkich ilości danych oraz wciąż rosnących możliwości ich przetwarzania rodzą się potrzeby czynienia dalszych kroków w kierunku nauk ścisłych, a zwłaszcza - rzecz jasna - informatyki. Doprowadziło to do rozwoju lingwistyki komputerowej, do budowy słowników elektronicznych oraz oprogramowań do przekładu (jeszcze nie zupełnie) automatycznego. Jednakże, żeby informatyka stać się mogła szansą dla lingwistyki, potrzebna jest bliższa współpraca między językoznawcami a informatykami. Zatem w drugiej części niniejszego artykułu opisane zostały dwa przykłady badań zarówno teoretycznych jak i z zakresu metodologii, które świadczą o próbach stworzenia nowego podejścia do zagadnień lingwistycznych, jakimi są pojęcia tak ważne, jak predykcja, topikalizacja i tym podobne, ale także o stosowaniu algorytmów

¹ Autor dziękuje za cenne rady i wskazówki prof. Bożennie Bojar (Uniwersytet Warszawski), prof. Radosławowi Katarzyniakowi (Politechnika Wrocławska) oraz dr. Pawłowi Stacewiczowi (Politechnika Warszawska), a także recenzentowi prof. Witoldowi Marciszewskiemu (Uniwersytet Warszawski). Winę za błędy i niedociągnięcia, jakie pomimo wszelkich starań, pozostały – być może – w niniejszym rozdziale autor ponosi osobiście, za co z góry pragnie przeprosić Czytających.

wypracowanych w ważniejszych działach eksploracji danych (data mining) celem ustanowienia nowych procedur dla rzeczowych badań interdyscyplinarnych wiodących, poprzez innowacyjne metody *interaktywne* (człowiek-komputer), do utworzenia w przyszłości lingwistyki prawdziwie *kompleksowej*.

Słowa kluczowe

lingwistyka strukturalna, lingwistyka informatyczna, lingwistyka interaktywna, skupianie uwagi w wypowiedzeniu, ontologia języka, semantyka, pragmatyka, orzekanie, metainformacja

6.1 Wstęp

Informatyka jest swoistym ucieleśnieniem czy raczej mechanizacją matematyki, bez której żadna metoda badawcza nie może dzisiaj zasługiwać na miano nauki, i to niezależnie od stopnia pożądanej ścisłości w prowadzonych badaniach. Stan aktualny lingwistyki strukturalnej jest niezupełnie zadowalający z punktu widzenia metodologii badań eksperymentalnych. Artykuł niniejszy reprezentuje stanowisko, według którego nauka jest jedna i zasadza się na matematyce, której część - od pół wieku - przybrała kształt informatyki, czyli inżynierii liczb, danych, informacji i wiedzy.

Dzisiaj zagadnieniem języka zajmują się eksperci wielu dziedzin badawczych, wśród których najważniejszymi są: neurologia mózgu, psychologia, teoria komunikacji, antropologia oraz estetyka. Zatem badania lingwistyczne - siłą rzeczy - winne być interdyscyplinarne, a ponadto eksploracyjne i interaktywne. Nie ulega wątpliwości, że to właśnie informatyka stanie się platformą, która umożliwi rozwój wiedzy w zakresie każdej z dziedzin badawczych, co - jednocześnie - uściśli wiedzę o tych dziedzinach oraz przyczyni się do ich systemowego opisanego, a często także i do symulowania poszczególnych mechanizmów, na jakie składają się zdolności językowe oraz funkcje mowy ludzkiej.

Dalsze unaukowanie lingwistyki jest potrzebą, o której pisze się coraz częściej. Nie tyle więc będzie mowa w tej pracy o przyczynach potrzeby unaukowania poszukiwań wiedzy o języku i językach, ile o nowych metodach naukowych i narzędziach badawczych. W Polsce na ten temat pisali m.in. A. Bogusławski (1986), W. Lapis (2003), J. Bańcerowski (2006), A. Pawłowski (2006) i D. Zielińska (2010). Zaś o potrzebie i sposobności stosowania narzędzi informatycznych interaktywnych w polonistyce pisała H. Włodarczyk (2010). Oprócz tego Bańcerowski J. (2006) opublikował także ważny artykuł w j. angielskim o tendencjach formalizowania oraz matematyzacji w lingwistyce XX stulecia.

6.2 Lingwistyka w dobie przemian

Pośród nauk humanistycznych i społecznych na przełomie XIX i XX stulecia w lingwistyce - jako najpierwszej - pojawił się paradygmat strukturalny (powstały w dużej mierze pod wpływem rozwoju chemii). Wkrótce też lingwistyka strukturalna stała się nauką "modną", co przyczyniło się do przyznania jej wiodącej roli wśród nauk takich jak socjologia, antropologia, psychologia poznawcza, teoria komunikacji oraz literatura i estetyka. Jednakże, ulegając wpływom metod opracowywanych w laboratoriach nauk komputerowych, w drugiej połowie XX stulecia "moda" przeobraziła się w szersze zainteresowanie językiem, jego złożoną budową oraz licznymi funkcjami, które można by usiłować poddawać matematycznej formalizacji. To właśnie w

atmosferze badań nad językami formalnymi (jakimi są m.in. sztuczne języki wymyślone dla celów programowania pierwszych elektronowych maszyn liczących) Chomsky N. (1956) opracował swoją teorię składni języków naturalnych², którą w niedługim czasie zdołał zainteresować światową wspólnotę lingwistów. Ujawniły się wtedy dwa przeciwstawne kierunki w lingwistyce strukturalnej³: obok deskryptywnego (indukcyjnego) powstał generatywny (dedukcyjny). Podczas gdy “deskryptywiści” zajmowali się różnorodnością materii językowej, zbieraniem i opisywaniem danych, “generatywiści” na podstawie domniemanej jednorodności formy językowej budowali gramatyki formalne. Sytuacja ta była przede wszystkim wynikiem różnicy w definiowaniu języka.

W lingwistyce *deskryptywnej* (opisowej) język - to zbiór wyrazów, analiza których prowadzi do poznania jego natury. Zasadniczą czynnością w deskrypcji języka jest wyjaśnianie (*explanation*).

$$J = (W, J)$$

Język - to zbiór wyrazów W otrzymany w wyniku analizy zbioru zdań należących do języka J .

W lingwistyce *generatywnej* język jest definiowany jako zbiór reguł, których znajomość pozwala na tworzenie wszystkich jego wyrażen. Model języka winien posiadać zdolność do *przewidywania* (*prediction*).

$$J = (W, G)$$

Język - to zbiór zdań utworzonych z wyrazów W wedle reguł gramatyki G .

Właściwy przełom w nauce o języku dokonał się jednakże nieco później, a mianowicie kiedy informatycy zaczęli opracowywać algorytmy do przekładu z języka na język. Okazało się wtedy, że oprócz formalnych teorii gramatyk generatywnych trzeba było wypróbować wiele innych podejść⁴. W wyniku badań laboratoryjnych⁵ z technik przetwarzania języków naturalnych wyrosła lingwistyka informatyczna (zwana także lingwistyką komputerową) mająca na celu zarówno analizę jak i syntezę wyrażen języków naturalnych. Jednakże pomimo wielu przydatnych zastosowań rozwiązań teoretycznych, badania prowadzone w tej dziedzinie polegały przede wszystkim na poszukiwaniu reguł gramatycznych z użyciem cech syntaktyczno-semantycznych, tzw. synsem, oraz wiedzy o preferencjach (statystycznych) w używaniu wyrażen językowych. W latach 90. w zasadzie zaprzestano rozwijania - przedtem niezwykle

² Warto zauważyć, że w tym samym czasie powstały (1) słynna dzisiaj “hierarchia języków formalnych” lub “hierarchia Chomskiego” a kilka lat później (2) dwa twierdzenia Chomskiego i Schützenbergera (1963) dotyczące (a) obliczania ilości słów w gramatykach bez-kontekstowych oraz (b) związku między algebrą a językami formalnymi.

³ Tak właśnie w roku 1987 Altman G. scharakteryzował podstawowe różnice pomiędzy badaniami prowadzonymi nad “formą” i “materią” językową.

⁴ Wśród nich (ok. 30-stu) przeróżne gramatyki takie jak LFG : Lexico-Functional Grammar (Bresnan 1982), FUG : Functional Unification Grammars (Key, 1983), HG : Head Grammars (Pollard, 1984), HPSG : Head-driven Phrase Structure Grammar (Pollard, 1985), GPSG : Generalized Phrase Structure Grammar (Gazdar et al., 1985), UCG : Unification Categorical Grammar (Uszkoreit, 1986), TAG : Tree-Adjoining Grammar (Joshi, 1987) itp.

⁵ Nie wszystkie z nich miały praktyczny charakter. Wystarczy wspomnieć Gramatykę Montague'a, której *raison d'être* był przecież aspekt tylko teoretyczny.

intensywnego - nurtu badań nad teoriami gramatyk formalnych, natomiast informatycy - w miarę możliwości, jak na przykład w Japonii - przenieśli swoje badania na różne uczelnie humanistyczne z programami nauczania nowej dziedziny lingwistycznej znanej pod nazwą przetwarzanie języków naturalnych (Natural Language Processing). Ponadto technikami przetwarzania języków naturalnych zainteresowali się poniekąd lingwiści będący wykładowcami uniwersyteckimi, którzy wprowadzili je do własnych programów nauczania. Zachowania te świadczyły o tym, że ówczesna inżynieria językowa osiągnęła swoje apogeum z powodów tak teoretycznych jak i sprzętowych, i że nabyte doświadczenia przyniosły świadomość konieczności przeprowadzenia dogłębnych badań nad językową “materią”.

Tablica Nr 1

	Lingwistyka strukturalna	Lingwistyka informatyczna
FORMA (Struktury)	Lingwistyka teoretyczna (transformacyjna, generatywna i aplikatywna)	Przetwarzanie języków naturalnych (Gramatyki leksykalno-funkcyjne, unifikacyjne i logiczne)
MATERIA (Dane)	Lingwistyka typologiczna (deskryptywna i korpusowa)	Technologie języków ludzkich (Zarządzanie bazami danych tekstowych, eksploracja tekstów i Lingwistyka interaktywna)

Należy także wspomnieć o istnieniu czysto matematycznego paradygmatu rozwijającego się równoległe zarówno z lingwistyką strukturalną jak i informatyczną. Do paradygmatu tego należą różne odmiany (podaję tylko inicjatorów) gramatyk kategoryalnych (Leśniewski, S. (1886-1939), Ajdukiewicz, K. (1890-1963), Lambek, J. - 1958, Steedman, M. - 2000) oraz stratyfikacyjnych (Lamb, S. - 1958). Gramatyka Montague'a jest jedną z ważniejszych odmian gramatyki kategoryalnej. Jej autor, logik Montague, R. (1970), postuluje wręcz, że języki naturalne należy traktować tak samo jak języki formalne. Jest to jednakże stanowisko dość kontrowersyjne, zwłaszcza w zaproponowanej wersji. W Polsce do badań zgodnych z paradygmatem matematycznym można zaliczyć między innymi prace Pogonowskiego, J. (1988, 1991, 1993, 1997 i inne).

6.3 Od materii do modelu

Lingwistyka korpusowa polega na prowadzeniu analiz jednostek językowych tworzących teksty, jej prawozorem były listy i kartoteki (dzisiaj pliki komputerowe) wyrażen językowych tworzonych przez lingwistów strukturalistów. Te praktyki zostały zautomatyzowane przy użyciu *technologii języków ludzkich*, której zasadniczym celem jest budowanie słowników i tezaurusów, a w grę wchodzi dzisiaj najnowocześniejsze (częściowo zazębiające się ze sobą) dziedziny badawcze będące pozyskiwaniem wiedzy o językach zwane (a) **drażeniem tekstów** (text mining), wyrastającym przede wszystkim z technik zarządzania bazami danych tekstowych⁶ i tych procedur drażnienia danych (data mining), które służą do wydobywania i obróbki danych

⁶ W Polsce w IPI PAN powstaje Narodowy Korpus Języka Polskiego (Przepiórkowski, A. 2004, 2012)

surowych oraz (b) **lingwistyką interaktywną**⁷, wyrastającą z technik statystycznej analizy danych (data analysis), drążenia danych (data mining) oraz zautomatyzowanego odkrywania wiedzy (automated knowledge discovery). Drążeniem tekstów nie będę się tutaj zajmował, ponieważ istnieją po polsku prace opisujące dość szczegółowo tę dziedzinę (na przykład Piasecki, M. 2010), natomiast postaram się zasygnalizować na tym terenie wciąż jeszcze słabiej rozwiniętą część badań z zakresu drążenia danych (data mining), a mianowicie specyfikację danych oraz ewaluację wyników analiz danych w odniesieniu do pozyskiwanej wiedzy od lingwistów w procesie interakcji z programami komputerowymi. Choć wyniki pracy algorytmów na surowym materiale są ważne i ciekawe, to jednak często potwierdzają one jedynie to, co już lingwiści w jakiś sposób opisali. Zatem najważniejszym etapem eksploracji danych w badaniach nad językami (nie zaś tylko tekstami) podczas stosowania narzędzi drążenia danych jest ich opis polegający na przekształcaniu danych surowych na symbole formalne, a następnie poddawaniu wyników tych przekształceń połowicznie zautomatyzowanym analizom i całkowicie zautomatyzowanym ewaluacjom.

6.3.1 Lingwistyka interaktywna

Lingwistyka interaktywna nawiązuje do postulatów *analizy składnikowej* (componential analysis) jako metody na wyodrębnianie składników bezpośrednich w jednostkach składniowych. Analiza składnikowa wraz z analizą tematyczną (theme analysis) i analizą taksonomiczną (taxonomic analysis) stanowiła podwaliny dla takich działów lingwistyki strukturalnej jak teoria pola leksykalnego, gramatyka transformacyjna oraz różne modele semantyki generatywnej. Chociaż została ona niesłusznie krytycznie oceniona i zaniechana na łonie lingwistyki strukturalnej⁸, użycie cech, a dokładniej *struktur cech* (feature structures), do opisu gramatyk okazało się nieodzowne w przetwarzaniu języków naturalnych (dział lingwistyki informatycznej zajmujący się pisanem gramatyk i słowników dla celów budowanych interfejsów (sprzęgów) człowiek-maszyna, przekładu automatycznego z języka na język itp.).

I rzeczywiście, podstawą naukowych przedstawień wszelkich dziedzin są właśnie cechy (własności) wchodzących w ich zakres przedmiotów. Podobnie rzeczy się mają w lingwistyce, gdzie przedmiotami badań są cechy składników wyrażen językowych należących do poziomów/modułów różnego rodzaju⁹. Ponieważ wyrażenia są znakami, traktując znaki jako byty same w sobie można rozpatrywać nie tylko byty (składniki) świata, lecz również i znaki językowe i wszelkie inne nośniki treści semiotycznych jako zawierające odpowiednio *treści ontologiczne* i *semantyczne*. Wszystkie rodzaje gramatyk formalnych, jakie są po dziś dzień używane do implementowania programów tłumaczących z języka na język, polegają na specyfikowaniu reguł semantyczno-syntaktycznych przy pomocy struktur cech. Problemem okazały się nie tylko przyjmowane (w żaden sposób jeszcze nie sprawdzone) cechy, lecz także zbyt jednolita (nie modularna) struktura reprezentacji. Celem metody interaktywnej jest zatem spójny (a więc sprawdzony) wybór cech podobnie, jak to jest obecnie czynione w programie

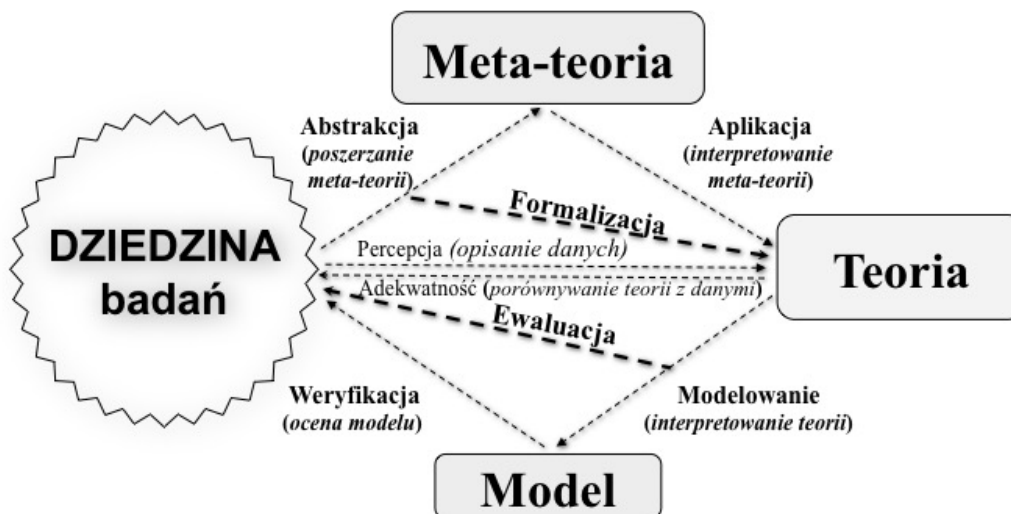
⁷ Termin ukuty przez autora niniejszego artykułu. Warto mimo to przytoczyć raczej wtedy wyjątkowy na świecie przykład bardzo owocnego zastosowania cech semantycznych do opisu ok. 200 czasowników ruchu, a była nim praca B. Bojar, 1979

⁸ Warto mimo to przytoczyć raczej wtedy wyjątkowy przykład bardzo szczegółowego, na łonie lingwistyki strukturalnej, zastosowania cech semantycznych do opisu ok. 200 czasowników ruchu, a była nim praca B. Bojar, 1979

⁹ Najczęściej wymieniane są cechy fonetyczne, prozodyczne, morfologiczne, składniowe, semantyczne i pragmatyczne.

gramatyki rozproszonej, (patrz 3.2). Badania polegają bowiem na procedurze iteratywnej (a więc w duchu Archimedesesa), mającej doprowadzić do uzyskania systemu wolnego od sprzeczności. Jest to niewątpliwie całkowita nowość w lingwistyce strukturalnej¹⁰, której dotychczasowe teorie nie miały żadnych szans ani na solidną formalizację pozyskiwanej wiedzy ani na jej ewaluację.

Schemat modelowania interaktywnego



Na schemacie modelowania interaktywnego uwidocznione zostały 4 rodzaje dziedzin: poza (a) **dziedziną badań** (fragmenty świata rzeczywistego tj. języki) znajduje się (b) **metateoria**¹¹ (użyteczne aksjomaty, abstrakcyjne pojęcia, struktury i operacje dotyczące wszystkich poziomów badanej dziedziny), (c) **teoria** będąca formalnym opisem fragmentu świata, która powstaje w wyniku przeprowadzonych badań oraz (d) **modele** (gramatyki, tezaury, leksykony itp.). Celem takiej procedury badawczej jest pozyskiwanie wiedzy poprzez budowanie systemów *reguł inferencji logicznych*, “kontekstów” (zerojedynkowych) i systemów *informacyjnych (wielowartościowych)* w drodze - w tej samej kolejności - operacjonalizacji systemów *doradczych (expert systems)* opartych na języku programowania *Prolog* (Colmerauer, A. 1970), systemów *decyzyjnych* (Pawlak, Z. 1981 i 1987) czy też logiki *kontekstowej* (Wille, R. 1982, Wolff, K. E. 2010). Dodajmy jeszcze, iż stosując logikę systemów rozproszonych¹² (Barwise, J. K. & Seligman, J., 1997) można wydzielić w tym iteratywnym procesie badawczym dwie abstrakcyjne sieci rozproszone, dzięki którym **formalizacja** i **ewaluacja** jawią się jako powiązane ze sobą komutacje przepływu informacji. Aspekt interaktywny badań polega zatem na tworzeniu meta-teorii i budowaniu modeli oraz na stosowaniu *formalizacji* i *ewaluacji* w iteratywnym (eksperymentalnym) procesie badawczym.

¹⁰ A nawet i w lingwistyce informatycznej poza nielicznymi przypadkami takimi, na przykład, jak gramatyki logiczne pisane w językach programowania takich, jak Prolog, ponieważ język ten ze swojej natury jest systemem pozwalającym na rozwiązywanie problemów.

¹¹ Metateorię do procedury iteratywnej zaproponował lingwistom Pogonowski J. (1991) włączając jednocześnie dziedzinę badań nieformalnych (wynik konceptualizacji), co za tym idzie etap rekonstrukcji teorii nieformalnych. Dziedzina ta jednakże traci rację bytu ze względu na możliwość interakcji z komputerem

¹² por. Stacewicz, P. & Włodarczyk, A. 2010 i 2011.

6.3.2 Przykład: teoria meta-informacyjnych centrów uwagi w wypowiedzi

Teoria meta-informacyjnych centrów uwagi (MIC¹³) w wypowiedzi¹⁴ jest pierwszą próbą uogólnionego opisu kilku zagadnień, często dyskutowanych oddzielnie w licznych pracach lingwistycznych, a mianowicie: *szyku wyrazów w zdaniu*, *predykcji*, *topikalizacji* oraz *fokalizacji*. Opracowana i sprawdzona w zastosowaniu do opisu języków różnych typów, wraz z teorią semantyki asocjacyjnej, stanowi ona zaczątek dla programu "gramatyki rozproszonej". Z uwagi zarówno na swoje podłoże informatyczne jak i na jej interdyscyplinarne konwergencje teoria ta rozwinęła się w wyniku badań prowadzonych z zastosowaniem metody interaktywnej. A to dlatego, że w teorii MIC (a) semantyka wypowiedzi językowych traktowana jest jak **informacja**, co umożliwia (b) podanie nowej definicji predykatu (orzeczenia) opartej na pojęciu **meta-informacji**¹⁵ dla opisu języków naturalnych tak na tle formuł logiki klasycznej jak i na użytek informatyki.

Według teorii semantyki asocjacyjnej jądrem znaczenia jest zatem informacja dotycząca *sytuacji* zdefiniowanej relacyjnie jako złożenie ról *czynnych*, *biernych* i *pośrednich* należących do różnych poziomów hierarchii sytuacji ontologicznych. Sytuacje wyrażane przez wypowiedzi językowe są najczęściej relacjami **wirtualnymi**, tj. takimi, które implikują "stowarzyszone" z nimi relacje **aktualne**. W skład definicji sytuacji semantycznej wchodzi także pojęcie *zakotwiczenia* czaso-przestrzennego. Ponadto nieodłącznymi (również podstawowymi) składnikami sytuacji semantycznych są *aspekt* i *modalność*.

Podstawowe składniki informacyjne wypowiedzi i jej wartości:

Sytuacja semantyczna = (R, Z), wartości = {prawda, fałsz},

gdzie **R** - role, **Z** - zakotwiczenia czaso-przestrzenne.

Podstawowymi składnikami warstwy komunikatywnej języka są wykładniki meta-informacji. Właściwie wydzielone zostały dwa poziomy meta-informacji: (1) struktura predykcji (tj. mówienia o podmiocie i ew. dopełnieniu) oraz (2) struktura topikalizacji i fokalizacji (tj. mówienia dodatkowo o topiku i ew. fokusie). Sposób, w jaki wykładniki meta-informacji wskazują na elementy informacji można porównać do **enkapsulacji**¹⁶ w programowaniu, gdzie rozgranicza się pomiędzy *sprzęgiem* (częściami programu "widocznymi", "publicznymi") oraz *implementacją* (częściami programu "ukrytymi", "prywatnymi"). W

¹³ MIC – to akronim od *Meta-Informative Centering* (Włodarczyk A. & Włodarczyk H. 2008, 2011 i 2013).

¹⁴ W niniejszym artykule termin "wypowiedź" został użyty w znaczeniu "językowy wykładnik (objaw) aktu mowy", podczas gdy wyrazy "wypowiadanie" i "wypowiedzenie" oznaczają pojęcie "aktu mowy" odpowiednio jako klasy i instancji.

¹⁵ Metainformację po raz pierwszy zdefiniował formalnie polski cybernetyk H. Greniewski (1968) w ramach definicji kodu będącego uogólnieniem pojęcia język, natomiast do badań nad językiem naturalnym włączyła je B. Bojar (1972, patrz także niżej). W innym ujęciu - także cybernetycznym - M. Mazur (1970) twierdził, że meta-informacja pojawia się już na poziomie sygnału, a mianowicie z chwilą transformacji informacji na wyrażenie językowe (w wyniku czego – wg tej teorii - powstaje potrzeba upewnienia się przez odbiorcę o prawdziwości komunikatu językowego).

¹⁶ Pojęcie *enkapsulacji* zapożyczam z informatyki zarówno z (a) programowania obiektowego (choć tylko częściowo; chodzi mianowicie o tę część definicji, która mówi o łączeniu struktur – wraz z danymi i wykonywanymi na nich operacjami – w jednostki (pakiety) często zwane modułami oraz (b) sieciowych systemów komunikacyjnych (patrz protokoły przesyłania danych itp.).

lingwistyce rozgraniczenie to było rozważane na różne sposoby i pod różnymi nazwami, bodaj najtrafniejszą z nich (choć niekoniecznie najlepiej zdefiniowaną) jest prawdopodobnie para “eksplikatura/implikatura” (Sperber, D. & Wilson, D. - 1998).

Podstawowe składniki meta-informacyjne wypowiedzi i jej wartości:

Enkapsulacja (oprawa) pragmatyczna = (*W*, *O*), wartości = {“datum”, “novum”},

gdzie *W*- wyróżnienie (“o czym” jest mowa) i *O* - omówienie (“co” się o tym mówi).

Z punktu widzenia teorii MIC enkapsulacja (to nie tylko hermetyzacja) odgrywa kapitalną rolę w budowaniu wyrażen językowych, bowiem komunikaty językowe są "oprawione" w podstawowe moduły meta-informacyjne) oraz zawierają wyłącznie treści cząstkowe kodu wewnętrznego, treści ewidentne pozostając "domniemanie"), a więc zamykane w format wyrażen języka naturalnego. Zatem języki naturalne enkapsulują meta-informację (warstwa pragmatyki) i informację (warstwa semantyki) nadając im w ten sposób kształtu sekwencji *fraz podstawowych*, tj. pierwszorzędnych składników bezpośrednich wypowiedzi (major phrases). To właśnie w Polsce powstała pierwsza w historii lingwistyki ogólnej propozycja (Bojar, B. - m.in. 1972, 1978, 1979 i 1986), której myślą przewodnią było zaliczenie meta-informacji wyrażanej przez czasowniki takie, jak “komentować”, “potwierdzać”, “namawiać”, “słuchać” i kilkadziesiąt innych do pragmatyki języka. Mimo to fakt, że ani meta-informacja ani enkapsulacja treści (tak semantycznych jak i pragmatycznych) w wypowiedzi w ogóle nie zostały dostrzeżone, był powodem tego, iż w lingwistyce ogólnej powstało sporo nieporozumień¹⁷ w sprawie opisu podstawowych części składowych wypowiedzi, jakimi są podmiot i orzeczenie. Ta sama informacja semantyczna może być tak *wyrażana* jak i *orzekana* (a nawet - jak zobaczymy *omawiana*) czy też - tak w przypadku orzekania¹⁸ jak i ogólniej - *enkapsulowana* z różnych punktów widzenia (aspekt¹⁹) zależnie od wyróżnienia takiej czy innej z jej części składowych. Będąc cechą modularności, enkapsulacja jest pojęciem, które - w porównaniu z jednolitą strukturą drzewiastą - pozwala na lepszą analizę systemu języka.

Inną niesłychanie ważną problematyką języków naturalnych jest to, że tak pojęcia jak i jednostki językowe podlegają równoległym funkcjom **selekcji** i sekwencyjnym funkcjom **kombinacji**. Dowody tego rodzaju procesów umysłowych płyną z badań nad afazją (choroba polegająca na utracie mowy), ponieważ - jak to po raz pierwszy opisał Jakobson, R. (1956), następnie zaś Wisniewski, E. J. (1997), u pacjentów cierpiących na jedną z odmian tej choroby zanikają zdolności porównywania cech, zaś u innych pojawiają się zaburzenia w łączeniu wyrazów. Maxwell, J. A. i Miller, B. (2012) zaproponowali włączenie selekcji i kombinacji do podstawowych składników *metody jakościowej* analizy danych. I rzeczywiście, biorąc pod uwagę, że wiedza naukowa jest przechowywana w formie narracji, kategoryzacja i wszelkiego rodzaju analogie winny służyć systematycznie i jednocześnie tak do prowadzenia analiz jak i do

¹⁷ Wprawdzie chociaż opisy niektórych autorów (m.in. M. A. K. Halliday, 1967, Halliday, M. A. K. & Greaves, W. S. 2008, Peregrin, J. 2011) wyróżniały różne płaszczyzny w wypowiedzi, zagadnienie predykcji pozostawało nadal w dziedzinie prawdziwościowej (semantycznej).

¹⁸ Warto zauważyć, że enkapsulacja “najszersza” zachodzi w wypowiedzi na poziomie meta-informacji.

¹⁹ W trakcie badań interaktywnych wypracowana została całkiem nowa teoria kategorii aspektu. Podstawy tej teorii stanowi część teorii sytuacji semantycznej (Włodarczyk A. 2003, Włodarczyk H. 2009, Włodarczyk A. & Włodarczyk H. 2013).

zapisu wyników. Tak też należy postępować w przypadku badań prowadzonych z zastosowaniem metody lingwistyki interaktywnej (patrz poniżej).

6.4 U zbiegu nauk

Informatyka, będąc zarazem działem matematyki i najlepszym przykładem jej zastosowania przy użyciu automatów (maszyn liczących) czyli inżynierią, stanowi nie tyle nowy paradygmat w nauce, ile wręcz solidne podwaliny do naukowego podejścia do wszelkich zagadnień. Struktury wiedzy oraz treści, jakie użytkownicy języków naturalnych/ludzkich sobie przekazują, muszą być niesłychanie złożone skoro - jak się okazało - przekład automatyczny wciąż nastrocza informatyce tak poważne trudności. Metoda informatyczna poza pojęciem **struktury** wprowadziła do badań nad językiem **reprezentację danych i informacji** oraz zasady ich **przetwarzania** (operacje). Jest to wynik algorytmicznego podejścia do zagadnień lingwistycznych.

6.4.1 Metoda interdyscyplinarna

U progu XXI stulecia badania naukowe w dziedzinie informatyki, a w szczególności w dziale tak zwanej sztucznej inteligencji, zapowiadają wielki przełom w metodologii badań. Niesłychane możliwości zapisywania wiedzy w bazach danych oraz nowe funkcje komputerów imitujące zdolności ludzkiego intelektu umożliwią w ciągu najbliższej dekady jeszcze bliższą współpracę między badaczami reprezentującymi różne nauki. Stanie się to o tyle skuteczne, że człowiek w swojej interakcji z komputerami będzie mógł *myśleć z maszynami*. Nie ulega wątpliwości, że zawsze kiedy człowiek tworzył nowe narzędzia pracy, jej wydajność rosła.

Myślenie z maszynami technicznie nazywane jest zwykle eksploracją danych, która to dziedzina informatyki wykorzystuje przeróżne metody obliczeniowe począwszy od analizy baz danych, włączając logiki rozmyte, przybliżone i decyzyjne, analizy pojęć formalnych, logiki systemów rozproszonych, metody obliczania ziarnistego zwane inaczej metodami obliczania przy pomocy słów (computing with words), a kończąc na automatycznym odkrywaniu wiedzy (automated discovery).

Podejście interdyscyplinarne okazuje się koniecznością zawsze wtedy, kiedy w grę wchodzi dziedzina badawcza wymagająca teorii języka uwzględniającej większą liczbę zmiennych niż klasyczne teorie wyłącznie immanentne²⁰. Interdyscyplinarność w połączeniu z metodą interaktywną umożliwia budowanie modeli, które już są lub będą bardziej adekwatne niż poprzednie. Program gramatyki rozproszonej (wyrastającej z paradygmatu informatycznych systemów współbieżnych i rozproszonych) oraz istniejące już teorie skupiania uwagi (centering of attention) w wypowiedzi i w tekście²¹ jest wynikiem takiego podejścia.

Jednakże badania interdyscyplinarne mają sens tylko wtedy, kiedy są wykonywane z rzetelnym przygotowaniem oraz zaangażowaniem badaczy w odpowiednie dziedziny wiedzy. Jak trudna jest współpraca interdyscyplinarna wystarczy wspomnieć wysiłki mające na celu tworzenie sieci *neuropodobnych*, czyli sztucznych sieci. Już w pierwszym okresie badań nad takimi sieciami (badań inspirowanych neurobiologicznie) liczba parametrów branych pod uwagę w rozwiązaniach formalnych była znikoma w stosunku do - z roku na rok rosnącej - ilości odkrywanych własności komórek nerwowych oraz ich obopólnych połączeń (dendryty i

²⁰ Cf. F. de Saussure (1916 i 1991) i L. Hjelmslev (1928 i 1979)

²¹ Centering theory (B. J. Grosz 1981, B. J. Grosz et al. 1995).

synapsy). Dzisiaj do sieci neuropodobnych wprowadza się już nie tylko modularność (a więc rozproszenie struktur), lecz wręcz konstruuje się systemy rozproszone, na które składają się procesy tak sekwencyjne jak i współbieżne, pamięci dzielone i prywatne, a “agenci” czy “aktorzy” wymieniają między sobą komunikaty zarówno w trybie synchronicznym jak i asynchronicznym. Neurolodzy żywią jednak wielkie nadzieje na współpracę z informatykami w procesie odkrywania modułów zwanych “strefami zainteresowań” (region of interest - ROI) dzięki nowym możliwościom wglądu *in vivo* w procesy zachodzące w mózgu za pomocą skanerów fMRI²².

Można się spodziewać, że interdyscyplinarne badania naukowe prowadzone w oparciu o ontologię (w liczbie mnogiej) z zastosowaniem metod interaktywnych pozwolą uporać się z niedociągnięciami kognitywizmu²³. Za takie “niedociągnięcie” można uważać niechęć do ontologii już nawet w klasycznej lingwistyce strukturalnej, a w lingwistyce kognitywnej wręcz jest nim częściowe zaprzeczenie, o ile nie odrzucenie wszelkiego odniesienia do ontologii.

6.4.2 Rozpoznanie konwergencji jako przykład podejścia kompleksowego

Dziedziną teorii MIC jest koncentracja *składników wypowiedzi* wynikająca ze skupienia uwagi. Jest to całkowicie zgodne z komputerową teorią centrowania *składników tekstu* (elementów wypowiedzi w odniesieniu do tekstu, B. J. Grosz 1981, B. J. Grosz et al. 1995). Choć powstałe w różnych okresach czasu (*resp.* 1990. i 1980.) i w różnych krajach (Francja i USA), teorie te wzajemnie się uzupełniają. Obydwie odwołują się do psychologicznego pojęcia, jakim jest funkcja uwagi w mózgu/umyśle człowieka. Warto zauważyć, iż pojęcia i definicje przyjęte w teorii MIC okazały się przydatne w psycholingwistyce do modelowania ważnych bloków aktywacji różnych stref pamięci ludzkiego mózgu (Stachowiak, Franz J. - 2013).

Podział fraz odnoszących do centrów uwagi w wypowiedzi bazowej na *globalne* i *lokalne* znalazł potwierdzenie w teorii uwagi Oberauera, K. (2002), neurologa specjalizującego się w badaniu struktur pamięci. W tej teorii możliwe są maksymalnie dwa centra uwagi jednocześnie, i to nawet pod warunkiem, że każde z nich – w zakresie tej samej hierarchii – należy do innego poziomu pamięci. By nazwać spełniające ten warunek operacje selekcji, Oberauer również - i to niezależnie od autorów teorii MIC - używa pojęć “uwagi globalnej” i “lokalnej”.

Według teorii MIC ogromną większość języków, jaką stanowią języki *nominatywne* (nazywane także *akuzatywnymi*), w wypowiedziach bazowych cechuje domniemany związek podmiotu z *agentem czynnym*, tzw. “agensem”), natomiast pozostałe języki należące do niezbyt licznej grupy języków *ergatywnych* w wyrażeniach bazowych cechuje domniemany związek podmiotu z *agentem biernym*, tj. “pacjensem”). W neuropsychologii uwagi, rozróżnienie wprowadzone przez Austina, J. H. (2006 i 2013) pomiędzy uwagą *ego-centriczną* a *allo-centriczną* zdaje się świadczyć o właściwym potraktowaniu wspomnianych różnic między językami różnymi językami etnicznymi w teorii MIC, co w konsekwencji znakomicie przyczynia się do lepszego (bo o niebo prostszego od istniejących) wyjaśnienia różnic pomiędzy wyżej wspomnianymi klasami języków.

Osobliwością teorii MIC jest *systematyczny* opis struktur (a) predykcji (orzekania) oraz (b) topikalizacji i fokalizacji (omawiania), jakie cechują nakładające się na siebie dwie struktury

²² fMRI jest akronimem od “functional Magnetic Resonance Imaging”.

(http://en.wikipedia.org/wiki/Functional_magnetic_resonance_imaging)

²³ O krytyce kognitywizmu zob. A. Chmielecki (2013).

hierarchiczne. Biorąc pod uwagę tylko wypowiedzi bazowe (kanoniczne), ich 3 formalne składniki (S=Podmiot, O=Dopełnienie i V=Czasownik) można zasadniczo ustawić parami w 4 szykach: (a) SVO i SOV oraz (b) VOS i OVS. Warto zauważyć, z punktu widzenia zdrowego rozsądku dot. komunikowaniu informacji zaledwie szyki SVO i SOV wydają się *normalne*; szyki VOS i OVS można uważać za mało prawdopodobne z uwagi na końcową pozycję globalnego składnika S (podmiotu) wypowiedzi. Natomiast, w drzewach składników bezpośrednich, szyki “liści” (terminali) takie jak *OSV i *VSO są absolutnie wykluczone²⁴, ponieważ w przeciwnym wypadku fraza odnosząca się do S (podmiotu) musiałaby być *lokalną* (a z definicji jest *globalną*). Zasadniczość szyków wyrazów **SOV** lub **SVO** potwierdzają dane pochodzące z lingwistycznej bazy danych WALS²⁵, z której wynika, że szyki te są charakterystyczne dla 2/3 z dotychczas opisanych 1377 języków świata.

6.5 Zakończenie

Mimo wielu lokalnych prób i “pobożnych” życzeń niektórych teoretyków języka, lingwistyka dzisiaj nadal często jest jeszcze dziedziną wiedzy co najwyżej “naukową”²⁶. Nie ulega więc wątpliwości, że zachodzi potrzeba szerszej debaty na temat jej dalszego unaukowania. Należy się spodziewać, że w niedalekiej przyszłości powstanie lingwistyka *kompleksowa*²⁷ (integral linguistics), która będzie nauką interdyscyplinarną opartą na matematyce oraz inżynierii informacji i wiedzy.

Teoria MIC została przytoczona w tym artykule z dwóch powodów: (1) by dać przykład zastosowania nad językiem naukowej metody polegającej na informatycznej pętli interaktywnej oraz - szerzej - (2) by wskazać na konieczność głębszego zrozumienia podstawowych struktur wypowiedzi ze względu na centralną rolę języka naturalnego we wszelkim dyskursie naukowym, co znalazło potwierdzenie w zastosowaniu metod statystycznych (tak za pomocą grupowania cech i obiektów) jak i automatycznego klasyfikowania obiektów, a nawet przy pomocy drażenia danych oraz informatycznych metod modelowania.

Warto też nadmienić, że wspomniane podejście interaktywne do pozyskiwania wiedzy teoretycznie ma klasyczne podłoże badawcze, które jest *analizyczne*. Przykładem tego niech będą najczęściej stosowane metody drażenia danych (data mining). Jednakże obecnie pośród metod, które imitują procesy biologiczne znane jako algorytmy genetyczne w paradygmacie badań ewolucyjnych, można znaleźć także bardzo oryginalne propozycje *syntetycznych* metod badawczych. Do takowych zaliczyć można z pewnością takie procesy neuropodobne, które powodują, że sztuczne sieci neuronowe²⁸ mogą się uczyć lub być nauczane. Pomimo, że - jak się wydaje - poglądy w tej dziedzinie są nie tylko przeciwstawne, a - jak w przypadku H. Nakashimy (2008) - nawet wręcz antagonistyczne, stanowią propozycję stosowania programowania ewolucyjnego jako **interaktywnej** metody badawczej.

²⁴ Nic też dziwnego, że w rozumowaniu dotyczącym “złożeń wywróconych” (“upside-down compositions”, Tellier, I. 2000), szyki wyrazów VSO i OSV nie mogą być wygenerowane w wyniku złożenia logicznego, ale muszą być uzupełnione przez rachunek lambda w celu linearyzacji fraz.

²⁵ Akronim od The World Atlas of Language Structures Online. <http://wals.info/feature/81A>

²⁶ Słownik Języka Polskiego PWN definiuje termin “naukowy” jako: **żart.** “silący się na naukowość”. Określenie to dotyczy przede wszystkim współczesnych “nauk o języku” (language sciences), nie dotyczy natomiast przykładowo komputerowego przetwarzania mowy (speech processing).

²⁷ W Polsce zob. Mazurkiewicz-Sokołowska, J. (2010).

²⁸ Stacewicz, P. (2015) proponuje nawet połączyć algorytmy genetyczne z sieciami neuronowymi.

Bibliografia

- Altman, G. (1987) "The Levels of Linguistic Investigation", *Theoretical Linguistics*, vol. 14: edited by H. Schnelle, W. de Guyter, Berlin - New York: 227-239.
- Austin, J. H. (2000) "Consciousness evolves when the self dissolves". *Journal of Consciousness Studies*, Volume 7, Numbers 11-12, Imprint Academic: 209-230.
- Austin, J. H. (2013) "Zen and the brain: mutually illuminating topics", *Journal Frontiers in Psychology - Consciousness Research*, Vol. 4, Art. 784: (www.frontiersin.org)
- Bańczerowski, J. (2006) "The axiomatic method in 20th century European linguistics", *History of the Language Sciences - An International Handbook on the Evolution of the Study of Language from the Beginnings to the Present*, Auroux, S., Koerner, E.F.K., Niederehe, H.-J., Versteegh, K. (eds), *Handbooks of Linguistics and Communication Science*, vol. 18 (3), Walter de Gruyter & Co. : 2007-2025.
- Barwise, J. K. & Seligman, J. (1997) *Information Flow, The Logic of Distributed Systems*. Cambridge, New-York : Cambridge University Press : 274.
- Bogusławski, A. (1986) "O pojęciu wyjaśniania i o wyjaśnianiu w lingwistyce", *Biuletyn PTJ*: 40 (45-51).
- Bojar, B. (1972) "Elementy metainformacji w tekstach języka naturalnego", *Prace filologiczne*, t. XXIII, Warszawa.
- Bojar, B. (1978) "Polskie czasowniki dotyczące procesów informacyjnych (elementy metainformacji w tekstach języka naturalnego)". *Studia językoznawcze. Streszczenia prac doktorskich III*. Ossolineum, Wrocław.
- Bojar, B. (1979) *Opis semantyczny czasownikołów ruchu oraz pojęć związanych z ruchem, Rozprawy Uniwersytetu Warszawskiego 168*, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa: 258.
- Bojar, B. (1986) "Metainformacja jako przedmiot badań pragmatyki języka naturalnego". *Acta Philologica*, nr 15.
- Bojar, B. (1991) "Metainformacja w języku naturalnym", *Words are Physicians for an Ailing Mind*, ed. by Grochowski, M., Weiss, D. and Sagners, *Slavistische Sammlung*, Vol. 17, Verlag Otto SAGNER, München: 93-99.
- Chmielecki, A. (2013), *Podstawy Psychoniki - ku alternatywie dla Cognitive Science*, IFiS PAN, Warszawa: 335 (37-81)
- Chomsky, N. (1956). "Three models for the description of language". *IRE Transactions on Information Theory* (2): 113–124.
- Chomsky, N. & Schützenberger, M.-P. (1963) "The Algebraic Theory of Context-Free Languages", in *Computer Programming and Formal Systems*, P. Braffort and D. Hirschberg (eds.), North Holland: 118-161.
- Colmerauer, A. (1970) "Les systèmes-q ou un formalisme pour analyser et synthétiser des phrases sur ordinateur", publication interne no 43, septembre 1970, Département d'informatique, Université de Montréal, republié dans *T.A.L.*, 1992 (1-2): 105-148.
- Greniewski, H., 1968, "Język nauki", *Zagadnienia Naukoznawstwa*, tom IV, zeszyt 1(13), PAN, Warszawa: 24-66.

- Grosz, B. J. (1981). "Focusing and description in natural language dialogues", *Elements of Discourse Understanding*, ed. by A. Joshi, B. Webber, and I. Sag. Cambridge University Press : 85-105.
- Grosz, B. J., Joshi, A. K., and Weinstein, S. (1995). "Centering: A framework for modeling the local coherence of discourse", *Computational Linguistics*, 21 (2).
- Hjelmslev, L. T. (1928) *Principes de grammaire générale*, wyd. Bianco Lundo, Kopenhaga. (przekład na j. polski: *Prolegomena do teorii języka, Językoznawstwo strukturalne. Wybór tekstów*, H. Kurkowska, A. Weinsberg (red.), Warszawa 1979).
- Jakobson, R. & Halle, M. (1956) "Two aspects of language and two types of aphasic disturbances", *Fundamentals of Language*, Roman Jakobson, The Hague, Mouton. (reprinted in: (a) Selected Writings - II: *Word and Language*, Mouton, The Hague – Paris 1971 and (b) *Essais de Linguistique Générale*, Éditions de Minuit, Paris 1963)
- Lamb, S. (1958) *A Grammar of Mono*. PhD. Dissertation. Berkeley: 391.
- Lambek, J. (1958), The mathematics of sentence structure, *Amer. Math. Monthly*, 65: 154–170.
- Lapis, W. (2003) "Jak się parać lingwistyką ? Formalna analiza podejść wieloaspektowych", *Investigationes Linguisticae*, vol. IX, Poznań.
- Halliday, M. A. K. 1967. Notes on Transitivity and Theme in English (Part 2). *Journal of Linguistics* 3 : 206.
- Halliday, M. A. K. & Greaves, W. S. (2008) *Intonation in the Grammar of English*, Equinox London.
- Maxwel, J. A. i Miller, B. (2012) "Real and Virtual Relationships in Qualitative Data Analysis (with Barbara Miller)", in Maxwell, J. A.: *A Realist Approach for Qualitative Research*, Part II: Chapter 7, George Mason University, Sage Publications.
- Mazur, M. (1970) *Jakościowa teoria informacji*, Wydawnictwa Naukowo-Techniczne, Warszawa: 224.
- Mazurkiewicz-Sokołowska, J. (2010) *Lingwistyka Mentalna w Zarysie – O zdolności językowej w ujęciu integrującym*, Universitas, TA)WPN Kraków : 261.
- Montague, R. (1970) "English as a Formal Language", *Linguaggi nella società e nella tecnica*, Bruno Visentini (ed.), Mailand: 189–223.
- Nakashima, H. (2008), "Methodology and a discipline for synthetic research — What is Synthesiology ?", in *Synthesiology*, English edition, Vol. 1 (2008) No. 4: 282-290.
- Oberauer, K. (2002) "Access to Information in Working Memory: Exploring the Focus of Attention", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2002, Vol. 28, No. 3: 411-421.
- Pawlak, Z. (1981) "Information Systems – Theoretical foundations". *Information Systems*, 6 (3): 205–218.
- Pawlak, Z. (1982) "Rough sets", *International Journal of Computer and Information Science*, vol. 11 (5): 341-356.
- Pawlak, Z. (1987) "O Analizie pojęć", in: *Od Kodu do kod*, A. Bogusławski & B. Bojar (red.): 249-252.
- Pawlak, Z. (1991) "Rough Sets". *Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht: 237.
- Pawłowski, A. (2006) "O stosowalności metod ścisłych w badaniach języka", in: *Efekt motyla - Humanisci wobec teorii chaosu*, Heck, D. i Bakula, K. (red.), Wydawnictwo

- Uniwersytetu Wrocławskiego, Wrocław: 92–99.
- Peregrin, J. (2011) “There is no such Thing as Predication”, *Conceptus - Zeitschrift für Philosophie*. Year 40 (97): 29-51.
- Piasecki, M. (2010) “Automatyczne wydobywanie wiedzy o semantyce języka naturalnego z korpusów tekstu”, *Metodologie językoznawstwa (Filozoficzne i empiryczne problemy w analizie języka)*, Stalmaszczyk Piotr (red.), Wydawnictwo Uniwersytetu Łódzkiego, Łódź: 143-181
- Pogonowski, J. (1988) *Matematyczny Model Analizy Lingwistycznej*, Working Papers series 29, UAM, Poznań: 135.
- Pogonowski, J. (1991) *Hierarchiczne Analizy Języka*, seria Językoznawstwo, Nr 14, Uniwersytet im. A. Mickiewicza, Poznań: 64.
- Pogonowski, J. (1993) *Linguistics Oppositions*, Wydawnictwo Naukowe Uniwersytetu im. A. Mickiewicza UAM, Poznań: 135.
- Pogonowski, J. (1997) “Przestrzenie Podobieństwa i Opozycji”, *Sklonność metafizyczna. Bogusławowi Wolniewiczowi w darze*. Wydział Filozofii i Socjologii Uniwersytetu Warszawskiego, Warszawa: 83–95.
- Przepiórkowski, A. (2004) *Korpus IPI PAN - Wstępna wersja*, Instytut Podstaw Informatyki, Warszawa.
- Przepiórkowski, A., Bańko, Górski, R. L. i Lewandowska-Tomaszczyk B. [eds.] (2012) *Narodowy Korpus Języka Polskiego*, Praca zbiorowa, Wydawnictwo Naukowe PWN SA: 331.
- Saussure (de), F. (1916) *Cours de linguistique générale*, publié par Ch. Bally, A. Sechehaye avec la collaboration de A. Riedlinger, Bally (w przekładzie na język polski: *Kurs językoznawstwa ogólnego*, Wydawnictwo Naukowe PWN, Warszawa 1991).
- Sperber, D. & Wilson, D. 1998. “Relevance and irony”, *Relevance Theory: Applications and Implications*, R. Carston & S. Uchida (eds.), Amsterdam: John Benjamins. 283-93.
- Stacewicz, P. & Włodarczyk A. (2010) “Modeling in the Context of Computer Science - a Methodological Approach”, *Journal Studies in Logic, Grammar and Rhetoric*, Philosophical Trends in the 17th Century from the Modern Perspective, a print and electronic journal, Halina Świączkowska (ed.), vol. 20 (33) : 155-179.
- Stacewicz, P. & Włodarczyk A. (2011) “O Modelowaniu Informatycznym - ze szczególnym odniesieniem do badań nad sztuczną inteligencją”, *Zagadnienia Naukoznawstwa*, 4 (190), Warszawa : 65-184.
- Stacewicz, P. (2015) “Evolutionary Schema of Modeling Based on Genetic Algorithms”, *Journal Studies in Logic, Grammar and Rhetoric*, vol. 40 (53) : 219-239.
- Stachowiak, F. J. (2013), “Semantic and episodic memory by reference to the ontological grounding of the old and new meta-informative status”, in [Włodarczyk, A. & Włodarczyk, H. eds. (2013)] : 103-119.
- Stachowiak, F. J. (2013), “Tracing the role of memory and attention for the meta-informative validation of utterances”, in [Włodarczyk, A. & Włodarczyk, H. eds. (2013)] : 121-141.
- Steedman, M. (2000), *The Syntactic Process*, The MIT Press.
- Tellier, I (2000), “Semantic-Driven Emergence of Syntax: The Principle of Compositionality upside-down”, *International Conference on the Evolution of Language (Evolang 2000)*, poster session, Paris : 220-224.

- Wille R. (1982) "Restructuring Lattice Theory: an Approach based on Hierarchies of Concepts", in: *Ordered Sets*, I. Rival (ed.), Reidel, Dordrecht-Boston: 445-470. Reprinted in: Ferré, S., Rudolph, S. (eds.): *Formal Concept Analysis. ICFCA 2009*. LNAI 5548. Springer, Heidelberg: 314-339.
- Wille, R. (2000) Contextual Logic Summary. In: Stumme, G. (ed.): *Working with Conceptual Structures: Contributions to ICCS*. Shaker-Verlag, Aachen : 265-276.
- Wille, R. (2004) Implicational Concept Graphs. In: Wolff, K.E. et al (eds.): *Conceptual Structures at Work*. LNAI 3127, Springer, Heidelberg: 52-61 :
- Wisniewski E., J. (1997) Understanding novel noun phrases. In *Proceedings of IWHIT 1997*, The University of Aizu: 97-101.
- Włodarczyk, A. (2003) "Les Cadres des situations sémantiques", *Études Cognitives / Studia Kognitywne* 5, SOW, PAN, Warsaw, p. 35-51, English translation: "Frames of Semantic Situations" in [Włodarczyk, A. & Włodarczyk, H. eds. (2013)] : 41-58.
- Włodarczyk, A. (2008) "Roles and Anchors of Semantic Situations", *Études cognitives / Studia kognitywne* 8, SOW, PAN, Warsaw, reprinted in [Włodarczyk, A. & Włodarczyk, H. eds. (2013)] : 3-20.
- Włodarczyk, A. (2009) "Interactive Discovery of Ontological Knowledge for Modelling Language Resources - prolegomena", *Representing Semantics in Digital Lexicography*, Proceedings of MONDILEX, the 4th Open Workshop, V. Koseska-Toszewa, L. Dimitrova and R. Roszko (eds), SOW, Warszawa : 44-55.
- Włodarczyk, A. & Włodarczyk, H. (2008) "Roles, Anchors and Other Things we Talk About : Associative Semantics and Meta-Informative Centering Theory", Series: "Mouton Series in Pragmatics", I. Kecskes (ed.), *Journal Intercultural Pragmatics*, Vol. 5. No. 3., Mouton - De Gruyters, Berlin/New York : 345-365.
- Włodarczyk, A. & Włodarczyk, H. (2011) "Information Centering: Subjecthood and Topicality", An International Workshop on Linguistics of BA and The 11th Korea-Japan Workshop on Linguistics and Language Processing, Waseda University, Tokyo.
- Włodarczyk, A. & Włodarczyk, H. eds. (2013) *Meta-informative Centering in Utterances - Between Semantics and Pragmatics*, John Benjamins Publishing Co.
- Włodarczyk, H. (2009) "From Ontological Attributes to Semantic Feature Structures — Experimental research on Aspect in Polish", *Cognitive studies – Études Cognitives*, vol. 9, SOW, Warszawa, 15-32.
- Włodarczyk, H. (2010) "Lingwistyka na polonistyce krajowej i zagranicznej w dobie filozofii informatyczno-logicznej", in *Journal LINGVARIA*, vol. 1 (7), Cracow.
- Wolff, K. E. (2010) "Temporal Relational Semantic Systems", *Conceptual Structures: From Information to Intelligence*, Lecture Notes in Computer Science, Volume 6208:165-180
- Zielińska, D. (2007) Proceduralny model języka. Językoznawstwo z pozycji teorii modeli nauk empirycznych. WUJ. Kraków.

Abstract

Linguistics in the second half of the 20th century was characterized by the pursuit of descriptions of languages that were as strict as possible. As a matter of fact, at the end of the 19th century, F. de Saussure searched for such a definition of language by which linguistics could become an autonomous discipline. However, it was under the influence of sciences such as mathematics and computer science that there arose the tendency to describe language

phenomena using formal methods and tools.

At the beginning of the 21st century, it seems that both structural and computational linguistics benefit from the achievements of computer science, which allows them to extract knowledge from multiple databases containing richer and richer material about individual languages. In view of the large amount of data and ever-increasing processing capabilities of computers, the need arises to take further steps towards science, and - especially of course - computer science. This has led to the development of computational linguistics, the construction of electronic dictionaries and (as yet not entirely) automatic translation software.

However, for computer science to become a real opportunity for linguistics, there is a clear need for closer cooperation between linguists and computer scientists. Thus, the second part of this article describes two examples of both theoretical research and methodology, which demonstrate attempts to create a new approach to some linguistic issues as important as predication, topicalization and the like, and also the application of algorithms developed in the major sectors of data mining in order to establish new procedures for interdisciplinary research methods via innovative *interactive* (human-computer) methods, in order to create a truly *integral* theory of language.

Key words

structural linguistics, computational linguistics, interactive linguistics, attention centering in utterance, language ontology, semantics, pragmatics, predication, meta-information