

Roman Urban  
Uniwersytet Wrocławski, Wrocław  
roman.f.urban@gmail.com

## SZKIC HISTORII LINGWISTYKI MATEMATYCZNEJ W ROSJI I ZSRR, ZE SZCZEGÓLNYM UWZGLĘDNIENIEM TŁUMACZENIA MASZYNOWEGO

**Słowa kluczowe:** lingwistyka matematyczna, tłumaczenie maszynowe, metody statystyczne w językoznawstwie, języki formalne, metodyka badań naukowych

**Keywords:** mathematical linguistics, machine translation, statistical methods in linguistics, formal languages, methodology of scientific research

### Wstęp

Lingwistyka matematyczna to nauka, która stosuje metody matematyczne do badań językoznawczych, a więc można ją ulokować na pograniczu matematyki i lingwistyki. Nie powinien wobec tego dziwić fakt, że matematycy odegrali niepoślednią rolę w jej rozwoju. Był nawet taki okres w badaniach nad tłumaczeniem maszynowym, kiedy matematycy i informatycy twierdzili, że nie potrzebują wiedzy językoznawczej, by tworzyć efektywne algorytmy tłumaczące, a Frederick Jelinek podobno miał stwierdzić: „Every time I fire a linguist from our project, the performance of our system gets better” (Hirst 2013: 718). Przyjmuje się, że lingwistyka matematyczna powstała w połowie XX w. w związku z próbą zastosowania komputerów do maszynowego tłumaczenia tekstów. Jednak już dużo wcześniej pojawiały się przypadki zastosowania matematyki w lingwistyce. Całkiem poważne zastosowania matematyki w badaniach językoznawczych spotykamy już na początku XIX w. Były to jednak próby odosobnione i daleko jeszcze było do określenia paradygmatu lingwistyki matematycznej. Dopiero problem tłumaczenia maszynowego okazał się niezwykle

stymulujący dla jej rozwoju i wytyczył kierunki badań na wiele dziesięcioleci. Z tego powodu wydaje się słuszne przyjęcie jako momentu powstania lingwistyki matematycznej daty rozpoczęcia w USA badań nad tłumaczeniem maszynowym, czyli roku 1954. Lata 50. XX w. były okresem zimnej wojny. Amerykanie potrzebowali systemu umożliwiającego automatyczne tłumaczenie tekstów (chodziło głównie o teksty naukowe) z języka swojego najpotężniejszego wroga, jakim był ZSRR. Pierwszy publiczny eksperyment, w którym komputer przełożył tekst rosyjski na język angielski, został przeprowadzony w 1954 r. w Nowym Jorku przez naukowców z Georgetown University i IBM. Konsekwencją podjęcia przez Stany Zjednoczone badań nad tłumaczeniem maszynowym, a w szczególności wspomnianego wyżej eksperymentu, było rozpoczęcie analogicznych badań w ZSRR.

Stworzona przez amerykańskiego matematyka i filozofa Norberta Wienera interdyscyplinarna nauka o ogólnych systemach oraz systemach samoregulujących i komunikacji, dająca formalizację sprzężenia zwrotnego, nazwana przez niego w 1948 r. cybernetyką, oraz teoria informacji (której podstawy zostały stworzone przez Claude'a E. Shannona także w 1948 r.) miały fundamentalne znaczenie dla rozwoju lingwistyki matematycznej, w szczególności zaś dla przetwarzania języków naturalnych i tłumaczenia maszynowego. Teoria informacji bazuje na teorii prawdopodobieństwa i nie byłaby możliwa bez aksjomatycznego ujęcia rachunku prawdopodobieństwa podanego przez jednego z najwybitniejszych matematyków XX w. i jednocześnie uczonego o niezwykle szerokich zainteresowaniach humanistycznych Andrieja N. Kołmogorowa w jego fundamentalnym dziele (wyd. niem. Kolmogoroff 1933, ros. Колмогоров 1936). W latach 60. XX w., gdy w humanistyce nastąpił zwrot w kierunku nauk ścisłych, Kołmogorow odegrał ważną rolę w procesie unaukowania humanistyki.

W polskiej literaturze naukowej brakuje opracowań na temat historii lingwistyki matematycznej w Rosji i ZSRR. Inaczej sytuacja wygląda w literaturze wydawanej w krajach zachodnich, choć i tam brak całościowego ujęcia problemu do czasów nam współczesnych. W tym miejscu trzeba przede wszystkim wymienić znakomitą monografię Ferenca Pappa (Papp 1966) obejmującą historię lingwistyki matematycznej w ZSRR do roku 1963 włącznie. Okres 1956–1962 jest bardzo szczegółowo omówiony w artykule Emmericha Keliha i Petra Grzybka (2005). W tym samym woluminie zamieszczony jest ich drugi artykuł traktujący o rozwoju kwantytatywnych metod w rosyjskim językoznawstwie i literaturoznawstwie, począwszy od lat 20. XIX w. do lat 30. XX w. (Grzybek, Kelih 2005). Wyczerpujące omówienie prac z zakresu rosyjskiego językoznawstwa matematycznego zawiera publikacja Sebastiana Kempgena (1995).

## 1. Pierwsze ślady lingwistyki matematycznej w Rosji i ZSRR

Rosyjskich zaczątków lingwistyki matematycznej dopatrzeć się możemy już w połowie XIX w.

W roku 1847 w czasopiśmie „Современник” rosyjski matematyk Wiktor Ja. Buniakowski opublikował artykuł, w którym po raz pierwszy zaproponował zastosowanie rachunku prawdopodobieństwa i statystyki w badaniach nad gramatyką i etymologią, a także w językoznawstwie porównawczym (Буныковский 1847), wspominając, że przekonanie o możliwości takich zastosowań czerpie z własnego doświadczenia, dokonał bowiem już pewnych prób w tym zakresie (zob. Sedlačiková 2012).

W 1855 r. w tym samym periodyku autor powieści *Что делать?* – Nikołaj G. Czernyszewski zamieścił artykuł, w którym ustosunkował się do pierwszego pełnego wydania dzieł Aleksandra Puszkina (Чернышевский 1855). Znajduje się w nim spostrzeżenie, że większość wierszy i poematów Puszkina jest napisanych w metrum dwójkowym (są to przede wszystkim jamby). W związku z tym autor zadał sobie pytanie, czy jamb nie jest najbardziej naturalnym metrum dla języka rosyjskiego. Licząc sylaby akcentowane i nieakcentowane z losowo wybranych tekstów, otrzymał stosunek 1 : 3. Na tej podstawie stwierdził, że jamb, który na 30 sylab wymaga 15 sylab akcentowanych, nie jest tak naturalny jak daktyl, amfibrach czy anapest, wymagające 10 sylab akcentowanych na 30. Później wracano jeszcze do tego problemu (Аверкиев 1893; Томашевский 1929), jednak to Czernyszewski był pionierem w tego rodzaju badaniach (Grzybek 2004: 92).

W 1883 r. Anton S. Budiłowicz w gramatyce języka staro-cerkiewno-słowiańskiego na podstawie zgromadzonych przez siebie materiałów podał statystykę występowania poszczególnych grafemów. Na przykład rozkład częstości samogłosek otrzymany przez Budiłowicza wygląda następująco: *i* – 20,4%; *ѣ* – 13,9%; *е* – 13,4%; *о* – 13,4%; *а* – 12,7%; *Ѣ* – 6,1%; *ѵ* – 6%. Dźwięki nosowe reprezentowane przez *Ѡ* oraz *ѡ* mają częstość występowania taką samą, równą 4,3%; *ѣѣ* – 3%, a najmniejszą *оѵ* – 2,5% (Будилович 1883: 67–68; zob. też Grzybek, Kelih 2004).

W 1913 r. Andriej A. Markow wygłosił na Wydziale Matematyczno-Fizycznym Rosyjskiej Akademii Nauk wykład, w którym naszkicował podstawy swojej nowej teorii dotyczącej pewnych ciągów zależnych zmiennych losowych, noszących obecnie nazwę **łańcuchów Markowa**. Teorię tę zilustrował na tekście *Eugeniusza Oniegina*, analizując prawdopodobieństwa wystąpień samogłosek i spółgłosek w tym utworze (Марков 1913).

Materiałem do badań było 20 000 pierwszych znaków *Oniegina* (z pominięciem znaków twardych i miękkich). Na podstawie tego materiału Markow obliczył prawdopodobieństwa wystąpienia samogłoski i spółgłoski oraz par samogłoska–samogłoska, samogłoska–spółgłoska itd. Zauważył, że zmienne losowe, będące wystąpieniami samogłoski bądź spółgłoski w danej pozycji, nie są niezależne. Zależą od tego, co występuje w pozycji poprzedzającej. Impulsem do badania występowa-

nia samogłosek w języku rosyjskim była dla Markowa korespondencja prowadzona ze statystykiem Aleksandrem A. Czuprowem (Sedlačíková 2012: 54).

Jan Baudouin de Courtenay, polski i rosyjski językoznawca, widział potrzebę uściślenia lingwistyki. Według niego miałyby się ono dokonać przez wprowadzenie metod matematycznych, a w szczególności metod statystycznych, których używał w swoich badaniach (Baudouin de Courtenay 1909; 1927/1984).

Uczniowie Baudouina, kazański – Wasilij A. Bogorodicki i petersburski – Lew W. Szczerba, byli pionierami stosowania metod eksperymentalnych w fonetyce oraz ścisłego i ilościowego podejścia do zagadnień języka (zob. Papp 1966: 17).

Jak wspomnieliśmy, impulsem do powstania lingwistyki matematycznej były próby maszynowego tłumaczenia tekstów. Tutaj także natrafiamy na Rosjanina – Piotra P. Trojanskiego<sup>1</sup>. Jak podaje John Hutchins:

Although we might trace the origins of ideas related to machine translation (MT) to 17<sup>th</sup>-century speculations about universal languages and mechanical dictionaries, it was not until the 20th century that the first practical suggestions could be made, in 1933 with two patents issued in France and Russia to Georges Artsrouni and Petr Trojanskij, respectively. Artsrouni's patent was for a general-purpose machine that could also function as a mechanical multilingual dictionary. Trojanskij's patent, also basically for a mechanical dictionary, went further with detailed proposals for coding and interpreting grammatical functions using 'universal' (Esperanto-based) symbols in a multilingual translation device (Hutchins 2006: 375).

## 2. Lingwistyka matematyczna w Związku Radzieckim

Przedstawione wyżej przykłady pokazują, że matematycy w Rosji interesowali się lingwistyką, a Markow wprowadzone przez siebie pojęcie szczególnego rodzaju procesu stochastycznego ilustrował, używając materiału literackiego. Łączenie matematyki i poezji w pracach matematyków będzie się zdarzać także później. Warto również podkreślić, że językoznawcy radzieccy bardzo często przejawiali duże zainteresowanie matematyką i doceniali jej rolę. Jak podaje Ferenc Papp (1966: 19), na IV Kongresie Sławistów w Moskwie w 1958 r. lingwista Piotr S. Kuzniecowa przyznał, że jego teoria przypadków zawdzięcza swą postać konsultacjom z matematykiem Andriejem N. Kołmogorowem. Z drugiej strony, jak podkreśla Aleksiej W. Gładki, wielu klasycznie wyedukowanych lingwistów miało mocno sceptyczne lub nawet wrogie poglądy na dokonujące się łączenie metod lingwistycznych i matematycznych:

Когда во второй половине 50-х годов некоторые молодые лингвисты задумались о применении математических методов для исследования структуры

1 Więcej informacji o Trojanskim – pionierze maszynowego tłumaczenia – i jego patentach zob. Бельская et al. 1959.

языка и начали сотрудничать с математиками, это вызвало у очень многих их коллег удивление и даже шок – ведь они с детства были убеждены, что гуманитарные науки, одной из которых является лингвистика, с математикой и другими «точными» науками не имеют и не могут иметь ничего общего (Гладкий 2005).

## 2.1. Srebrny wiek lingwistyki matematycznej

Lata 50. XX w. w lingwistyce radzieckiej Władimir A. Uspienski nazywa „srebrnym wiekiem” lingwistyki strukturalnej, stosowanej i matematycznej (Успенский 1998). Rajmund G. Piotrowski i Jurij Romanow wspominają tę początkową fazę badań jako „okres entuzjazmu i romantycznej nadziei” (Piotrowski, Romanov 1999: 105). Czas chruszczowowskiej odwilży sprzyjał naukowcom pracującym w dziedzinie lingwistyki matematycznej. Wcześniej, w czasach stalinowskich, cybernetyka, z którą lingwistyka matematyczna miała ścisłe związki, była uważana w Związku Radzieckim za „naukę burżuazyjną”, wobec czego jej uprawianie bez poparcia władz było utrudnione. Dla rozwoju cybernetyki kluczową rolę odegrali dwaj matematycy, wspomniani już wcześniej Andriej N. Kołmogorow oraz Aleksiej A. Liapunow<sup>2</sup>, a także radiotechnik Aksel I. Berg, który był dyrektorem utworzonego w 1955 r. Instytutu Radiotechniki i Elektroniki Akademii Nauk ZSRR (Институт радиотехники и электроники) i który miał też duży wkład w rozwój lingwistyki strukturalnej i sztucznej inteligencji.

Rozwój badań nad lingwistyką matematyczną był niezwykle dynamiczny, o czym może świadczyć liczba naukowców pracujących w tej dziedzinie. Pod koniec lat 50. XX w. w ZSRR było więcej naukowców pracujących nad tłumaczeniem maszynowym niż we wszystkich innych krajach razem wziętych (Harper 1960).

## 2.2. Tłumaczenie maszynowe

Począwszy od roku 1954 w największych ośrodkach naukowych ZSRR (Moskwa, Leningrad, Gorki (obecnie Niżny Nowogród), Kijów, Tbilisi, Erewań, Wilno (Piotrowski, Romanov 1999: 105)) powstały grupy pracujące nad tłumaczeniem maszynowym. Omówimy najważniejsze z nich.

---

2 O podejściu Liapunowa do problemu tłumaczenia maszynowego zob. Журавлёв 1980.

### 2.2.1 Moskwa

We wrześniu 1954 r. pojawiła się praca Dmitrija Ju. Panowa o amerykańskim systemie tłumaczącym (Панов 1954). Tę datę można uznać za początek badań nad tłumaczeniem maszynowym w ZSRR (więcej zob. Papp 1966; Roberts, Zarechnak 1974; Успенский 1998; Hutchins 2000, 2013).

Dmitrij Ju. Panow kierował grupą zajmującą się tłumaczeniem maszynowym w Instytucie Mechaniki Precyzyjnej i Techniki Obliczeniowej (Институт точной механики и вычислительной техники АН СССР – ИТМВТ). Zaangażował do prac nad maszynowym tłumaczeniem Izabellę K. Bielską. Już pod koniec 1955 r. Panow i Bielska przetestowali na komputerze pierwszy algorytm programu tłumaczącego z angielskiego na rosyjski. Władimir A. Uspienski ocenia ich podejście jako niezbyt udane, a jako przyczynę podaje odejście od metod strukturalnych i działanie w izolacji od lingwistów (Успенский 1998). Podsumowanie badań Panowa zob. Панов 1956a.

System tłumaczący skonstruowany przez grupę Panowa był ulepszeniem systemu opracowanego przez Georgetown University i IBM. Zawierał między innymi algorytmy pozwalające na morfologiczną analizę tekstu. Dokładne opisy można znaleźć w: Панов 1956a, b, c. Opis wyników uzyskanych do 1957 r. zob. Belska-ya 1957.

Drugą moskiewską grupę zajmującą się tłumaczeniem maszynowym utworzył Aleksiej A. Liapunow w Instytucie Matematyki im. Stieklowa (Математический институт им. В.А. Стеклова). Pierwszą publikacją był artykuł Liapunowa<sup>3</sup> i Olgi S. Kułaginy referujący amerykański eksperyment naukowców z Georgetown University (Ляпунов, Кулагина 1955).

Pod koniec 1954 r. O.S. Kułagina i Igor A. Mielczuk z Instytutu Językoznawstwa (Институт языкознания АН СССР) rozpoczęli pracę nad systemem tłumaczącym teksty matematyczne z francuskiego na rosyjski. Takie ograniczenie rodzaju tłumaczonego tekstu ułatwiało zadanie, teksty matematyczne bowiem charakteryzują się ograniczonym zasobem słów i utartymi zwrotami. Na początku 1956 r. algorytmy i specjalnie zbudowany słownik były gotowe do testów (Кулагина, Мельчук 1956; Кулагина 1958a).

W 1967 r. Kułagina zastosowała inne podejście do tego samego problemu. Użyła pewnych drzewiastych struktur odzwierciedlających zależności w zdaniach (Hutchins 2013).

Od 1955 r. Tatiana N. Mołosznaja pracowała nad tłumaczeniem z angielskiego na rosyjski tekstów matematycznych. Ponieważ w wypadku angielskiego mniej informacji można uzyskać ze struktury morfologicznej, zadanie wymagało innego podejścia, opartego bardziej na budowie syntaktycznej tekstu (Mološnaja 2000).

---

3 O wkładzie Liapunowa w rozwój tłumaczenia maszynowego zob. Кулагина 1977.

Mielczuk zajmował się także tłumaczeniem maszynowym z języka węgierskiego na rosyjski (Мельчук 1958a). Nie przekształcił jednak algorytmów w działający program komputerowy, doszedł bowiem do wniosku, że nierozwiązane kwestie wciąż pozostają i niezbędna jest dalsza analiza, w szczególności zaś odpowiednia reprezentacja znaczenia w tekście wyjściowym (Мельчук 1958b). Niepowodzenie na tym polu stało się dla Mielczuka bodźcem do trwających wiele lat (lata 60. i 70.) badań, których wynikiem jest semantyczny model języka znany pod nazwą *teoria SMYŚL⇔TEKST* (ros. *теория смысл ⇔ текст*, ang. *Meaning ⇔ Text theory*). Pełna nazwa teorii brzmi: *teoria lingwistycznych modeli SMYŚL⇔TEKST*. W jej powstaniu mieli swój udział także Aleksandr K. Żółkowski i Jurij D. Apresjan (zob. Мельчук 1974).

Teoria ta opisuje, w jaki sposób znaczenie zostaje przetransformowane w wypowiedź, czyli w tekst, i odwrotnie, jak z tekstu tworzone jest znaczenie. A zatem jest ona opisem pewnego rodzaju translatora. Przejście od znaczenia do tekstu (i odwrotnie) jest procesem wielostopniowym, sekwencyjnym. Na każdym stopniu tego procesu powstaje nowa reprezentacja wyjściowego znaczenia. Pierwszym stopniem jest reprezentacja semantyczna (w postaci grafu zależności semantycznych). W kolejnych stopniach powstają struktury syntaktyczne powierzchniowe i głębokie (w postaci struktur drzewiastych). Dalej – tworzy się linearna struktura morfologiczna, która może być zamieniona na reprezentację foniczną.

W związku z teorią SMYŚL⇔TEKST i jej zastosowaniem do tłumaczenia maszynowego pod kierunkiem Igora A. Mielczuka została zrealizowana jego koncepcja dotycząca budowy słownika semantyczno-kombinatorycznego (*толково-комбинаторный словарь*) (Мельчук, Жолковский 1984). Słownik zawiera opis słowa, jego znaczenie zapisane w odpowiednim metajęzyku oraz model rządu danego słowa. Słownik powstawał w latach 60. i 70. przy udziale grupy liczącej ponad 20 osób. O znaczeniu słownika semantyczno-kombinatorycznego dla algorytmów tłumaczenia maszynowego zob. też Way 1990.

Bazując na teorii SMYŚL⇔TEKST Mielczuka, Ju.D. Apresjan ze współpracownikami opracował system o nazwie ЕТАР (ЭТАП) tłumaczący z języka rosyjskiego na angielski i odwrotnie (Apresjan et al. 1992).

Kolejna moskiewska grupa, kierowana przez Wiktora Ju. Rozencwejga, działała w powołanym przez niego do życia Laboratorium Maszynowego Tłumaczenia przy Moskiewskim Pedagogicznym Instytucie Języków Obcych (Лаборатория машинного перевода при Московском государственном педагогическом институте иностранных языков им. Мориса Тореза). Rozencwejg uważał, że jeżeli chce się otrzymać systemy tłumaczące dające dobre przekłady, to badania nad tłumaczeniem maszynowym nie mogą pomijać semantyki.

Aleksandr K. Żółkowski, Nina N. Leontjewa i Ju. S. Martemjanow pracowali nad formalizacją znaczenia. W tym celu próbowali zdefiniować tak zwane *mnożniki semantyczne*, to znaczy pewne jednostki semantyczne, za pomocą których można by wyrażać bardziej skomplikowane treści (Жолковский et al. 1960, 1961).



Ciekawe podejście do problemu tłumaczenia maszynowego zaproponował Izaak I. Riewzin, robiąc użytek z wprowadzonego przez Lwa W. Szczerbę rozróżnienia gramatyki *aktywnej i pasywnej* (Ревзин 1958a). O pracach Rewzina będzie jeszcze mowa niżej. W tym miejscu odnotujmy tylko, że Riewzin i Rozencweyg są autorami podręcznika akademickiego *Основы общего и машинного перевода* (Ревзин, Розенцвейг 1964).

### 2.2.2. Leningrad

Nikołaj D. Andriejew kierował założoną przez siebie w 1955 r. grupą naukowców pracujących nad tłumaczeniem maszynowym w Instytucie Językoznawstwa Akademii Nauk ZSRR w Leningradzie (Институт языкознания АН СССР). Miała ona inne niż grupy moskiewskie podejście do tłumaczenia maszynowego. Wykorzystywała i rozwijała sztuczny język pośredni (metajęzyk zbudowany na bazie logiki formalnej) – tak zwaną *interlingwę* (Andreev 1967). Jeżeli problem dotyczy tłumaczenia z jednego języka na drugi, to interlingwa ma być językiem skonstruowanym w taki sposób, by zawierał wszystkie specyficzne własności gramatyczne języków, między którymi chcemy dokonać przekładu. Najpierw tłumaczono z języka wyjściowego na interlingwę, a potem z interlingwy na język docelowy.

### 2.2.3. ALPAC i zmierzch srebrnego wieku w lingwistyce matematycznej

W 1960 r. Yehoshua Bar-Hillel opublikował artykuł, w którym krytycznie wypowiedział się na temat powszechnie w tym czasie panującego przekonania o możliwości w pełni automatycznego tłumaczenia o wysokiej jakości (*fully automatic high quality translation* – FAHQT, Bar-Hillel 1960). Jego zdaniem FAHQT nie będzie nigdy możliwe i należy albo zadowolić się tłumaczeniem o niższej jakości, albo zaakceptować konieczność edycji wynikowego tekstu przez człowieka. W 1964 r. na prośbę organizacji finansujących badania nad tłumaczeniem maszynowym w USA National Science Foundation (NSA) powołała specjalną komisję Automatic Language Processing Advisory Committee (ALPAC) w celu zbadania faktycznego stanu rzeczy. Komisja w raporcie (ALPAC 1966) stwierdziła między innymi, że maszynowe tłumaczenie jest mało dokładne i dwa razy droższe od przekładu dokonywanego przez tłumacza (Hutchins 2010). Raport miał decydujący wpływ na znaczne zmniejszenie finansowania badań nad maszynowym tłumaczeniem w USA. Okazało się, że miał on także ważny wpływ na sytuację naukowców w ZSRR pracujących w dziedzinie tłumaczenia maszynowego, w szczególności na zmniejszenie finansowania ich badań. R.G. Piotrowski i Ju. Romanow stwierdzają bowiem:



But the main blow was dealt by the ALPAC Report. It is important to remind that, for the Soviet bureaucracy, an American opinion on scientific, economic or military problems was of paramount authority (Piotrowski, Romanov 1999: 106).

Dalej piszą, że raport miał fatalne konsekwencje dla wielu grup zajmujących się tłumaczeniem maszynowym. Rosyjska Akademia Nauk i ministerstwa odpowiedzialne za finansowanie badań wstrzymały dotacje. Decyzje te zostały przychylnie przyjęte przez lingwistów wykształconych tradycyjnie, a ich uprzedzenie do automatycznego tłumaczenia przybrało na sile. Sytuacja była na tyle trudna, że wielu naukowców, między innymi Nikołaj D. Andriejew, wycofało się z badań w tej dziedzinie (ibid.: 107). Skończyła się „romantyczna era” w tłumaczeniu maszynowym, oparta głównie na podejściu teoretycznym. Zaczął się okres dominacji pragmatycznego, inżynierskiego podejścia do problemu (ibid.; Roberts, Zarechnak 1974: 2841).

Piotrowski uważał, że większość grup naukowych pracujących nad maszynowym tłumaczeniem można było uchronić przed rozpadem, gdyby ówczesna sytuacja polityczna była normalna. Niestety tak nie było. Po entuzjastycznym okresie lat 50. maszynowe tłumaczenie

[...] became a coin of barter in the intrigues and rivalries between Committee for Science and Technology and the Academy of Sciences, between KGB and the Soviet Army, and between the Communist Party bureaucracy and the military-industrial structures (Piotrowski 2000: 236).

#### **2.2.4. Prozaiczne czasy – po raporcie ALPAC**

Jak to określili R.G. Piotrowski i Ju. Romanow, w latach 60. XX w. zaczyna się w Rosji „prozaiczny okres badań nad tłumaczeniem maszynowym”. Minęła już euforia lat 50., okazało się bowiem, że przez dziesięć lat mimo uzyskania wielu ważnych wyników teoretycznych nie udało się osiągnąć zadawalających wyników praktycznych. Prace nad tłumaczeniem maszynowym, choć w znacznie mniejszej skali, trwały jednak dalej.

##### **2.2.4.1. Wszechzwiązkowa Grupa „Statystyka mowy”**

Po raporcie ALPAC ważną postacią stał się Rajmund G. Piotrowski, który w latach 60. w Instytucie Pedagogicznym im. A.I. Hercena w Leningradzie (Ленинградский педагогический институт им. А.И. Герцена) zorganizował Laboratorium Lingwistyki Inżynierskiej (Лаборатория инженерной лингвистики). Prowadzono w nim badania statystyczne języka, konstruowano systemy automatycznego przetwarzania tekstów, w szczególności zaś pracowano nad tłumaczeniem maszynowym. W 1964 r. Piotrowski utworzył Wszechzwiązkową Grupę „Statystyka mowy” (Всесоюзная

группа „Статистика речи”), której oddziały znajdowały się we wszystkich większych miastach ZSRR, a później także w USA, Francji, Holandii, Szwajcarii, Izraelu i Afganistanie. Przez 30 lat pracowano nad problemami wymienionymi wyżej, a także zajmowano się lingwistycznymi aspektami sztucznej inteligencji. Na początku lat 80. „Statystyka mowy” opracowała system przetwarzania języka naturalnego, który był systemem maszynowego tłumaczenia z języka chińskiego na rosyjski i został użyty do przetwarzania tekstów przesyłanych z Chin. Był on częścią większego systemu śledzącego sytuację polityczną i ekonomiczną w ChRL (Piotrowski, Romanov 1999: 107; Hutchins 2013).

„Statystyka mowy” zaraz po powstaniu przyjęła jako cel swojej pracy skonstruowanie praktycznych systemów tłumaczących. Zaczęto badać statystyczne własności tekstów i wypowiedzi. Głównym teoretycznym osiągnięciem tego początkowego okresu było zidentyfikowanie przez R.G. Piotrowskiego trzech fundamentalnych niezgodności między językiem naturalnym a możliwościami przetwarzania go przez komputer (wyd. ros. Пиотровский 1975; wyd. niem. Piotrowski 1984). Oto te rozbieżności:

1) Język naturalny ma charakter ciągły i rozmyty (ang. *fuzzy*). Pojawiają się bowiem w nim na przykład wyrażenia typu *chłodno, zimno, ciepło, gorąco, blisko, daleko*, niemające ściśle zdefiniowanego znaczenia. Przy ilu stopniach Celsjusza możemy powiedzieć, że jest *ciepło*? Można tu co najwyżej podać jakiś przedział temperatur, który większość użytkowników języka zaakceptuje jako odpowiadający określeniu *ciepło*. Natomiast język komputera jest językiem operującym na dyskretnych (nierozmytych) wielkościach. Używa tradycyjnych (nierozmytych) zbiorów. Konieczne jest zatem wykorzystywanie w badaniach nad językiem teorii zbiorów rozmytych i logik rozmytych.

2) Język naturalny jest otwartą i dynamiczną (diachroniczną) strukturą, jest rozmytym komunikacyjnym systemem, podczas gdy charakter procesów zachodzących w komputerze jest statyczny (synchroniczny).

3) Wieloaspektowa natura ludzkich wypowiedzi (trzy sensory/znaczenia: nadawcy, odbiorcy i uniwersalny) nie może być przez komputer „zrozumiana”. Rozumienie komputera jest bowiem bardzo ograniczone i redukuje się do jednego możliwego znaczenia, jakie komputer jest w stanie wydobyć z wypowiedzi.

Te obserwacje wyznaczyły nowe podejście grupy do problemu tłumaczenia maszynowego (Piotrowski, Romanov 1999: 110–112).

W 1974 r. większość grup pracujących nad maszynowym tłumaczeniem została scentralizowana w powstałym Wszechzwiązkowym Centrum Przekładów Literatury i Dokumentacji Naukowo-Technicznej (Всесоюзный центр переводов научно-технической литературы и документации), kierowanym przez Jurija N. Marczuka.

Rozpad imperium radzieckiego pociągnął za sobą zmiany w działalności grupy. Brak centralnego finansowania spowodował, że grupy, które znalazły się na Ukrai-

nie, Białorusi i w Mołdawii, zaczęły zajmować się głównie działalnością komercyjną i sprzedają na europejskim rynku programów użytkowych służących do przetwarzania wyspecjalizowanych języków naturalnych – głównie biznesowych (ibid.: 108).

### 2.2.5. Komercyjne i użytkowe systemy tłumaczące w ZSRR

Po 10 latach od rozpoczęcia prac nad tłumaczeniem maszynowym pojawiły się i zaczęły być używane w ZSRR pierwsze działające systemy tłumaczące. Od 1964 r. główny organ zajmujący się przetwarzaniem informacji o patentach – Centralny Naukowo-Badawczy Instytut Informacji o Patentach (Центральный научно-исследовательский институт патентной информации – ЦНИИПИ) zaczął używać systemu tłumaczącego z języka angielskiego na rosyjski do przekładu tekstów patentów publikowanych przez Amerykańskie Biuro Patentów (US Patent Bureau) (Hutchins 2013). System został opracowany przez grupę Leonida G. Krawca (Кравец Василевский, Дубитская 1967).

W 1976 r. członkowie „Statystyki mowy” pod kierownictwem Kałdybaja B. Biek-tajewa i P.W. Sadczikowej w Szymkent w Kazachstanie opracowali system do przekładu z języka angielskiego tekstów patentów w dziedzinie chemii (Pevzner 2001: 85).

Ju. N. Marczuk, kierownik Wszechzwiązkowego Centrum Przekładów Literatury i Dokumentacji Naukowo-Technicznej, nadzorował w Moskwie prace nad trzema systemami tłumaczącymi: angielsko-rosyjskim AMPAR (АМПАР), niemiecko-rosyjskim NERPA (НЕРПА) i francusko-rosyjskim FRAP (ФРАП). Systemu AMPAR zaczęto używać w 1977 r. do wstępnego tłumaczenia tekstów technicznych. Program był używany przez Centrum, a także przez inne jednostki badawcze (Pevzner 2001: 85).

W 1991 r. powstała w Leningradzie opracowana przez Swietlanę Sokołową pierwsza wersja systemu tłumaczącego STYLUS. Był on rozwijany do 1997 r. (Sokolova 1997) i służył do tłumaczenia korespondencji biznesowej z języka angielskiego i niemieckiego na rosyjski oraz z rosyjskiego na angielski. W rok później dodano możliwość tłumaczenia między rosyjskim i francuskim, rosyjskim i włoskim, niemieckim i francuskim oraz angielskim i francuskim (Hutchins 2013). O systemach tłumaczących powstałych już po rozpadzie Związku Radzieckiego zob. Hutchins 2013.

### 2.3. Badania nad maszynowym przekładem a językoznawstwo ogólne

Badania, które zostały podjęte ze ściśle określonym celem umożliwienia maszynowego przekładu, przyniosły wiele interesujących wyników mających szersze znaczenie i wniosły istotny wkład do językoznawstwa ogólnego. W tym miejscu jako

jeden z możliwych przykładów należy wymienić teoriomnogościowy model języka Kułaginy (Кулагина 1958b, c), nad którym pracowali również A.A. Liapunow oraz I.I. Riewzin. Artykuł Kułaginy (Кулагина 1958b) był próbą ścisłego opisu części mowy, w szczególności autorka rozważała w nim podział leksemów na klasy ze względu na możliwość występowania w tych samych kontekstach. Problemem definicji części mowy zajmowali się także Władimir A. Uspieski (Успенский 1957) oraz I.I. Riewzin (Ревзин 1960, 1962).

Riewzin podał różnego rodzaju interpretacje podstawowego modelu Kułaginy w serii prac (Ревзин 1958b, 1959, 1960). Jego matematyczny model języka (Ревзин 1962) spotkał się z zainteresowaniem naukowców poza ZSRR, a monografia *Модели языка* doczekała się przekładu na język angielski (Revzin 1966).

Jeśli chodzi o szczegółowsze zagadnienia, to trzeba tu wspomnieć, że Roland L. Dobruszyn już w 1957 r. podejmował próby definiowania pojęć gramatycznych, stosując metody teorii mnogości (Добрушин 1957, 1958b). Z kolei problem rodzaju gramatycznego był badany z różnych punktów widzenia. I.A. Mielczuk badał zależność między rodzajem rzeczownika a jego końcówką w języku francuskim (Мельчук 1959), Andriej A. Zalizniak badał syntaktyczny aspekt rodzaju (Зализняк 1964), a Riewzin analizował rodzaj gramatyczny w ujęciu diachronicznym (Ревзин 1962).

Analiza syntaktyczna tekstu była w początkowym okresie prac nad tłumaczeniem maszynowym jednym z najważniejszych problemów. Algorytmy takiej analizy przedstawił Mielczuk (Мельчук 1962).

Aleksiej W. Gładki zajmował się między innymi językami i gramatykami formalnymi. Osiągnął wiele ważnych rezultatów w lingwistyce formalnej (inspirowanych badaniem języków naturalnych). O zastosowaniu metod matematycznych (języków i gramatyk formalnych) do badania języków naturalnych zob. Гладкий 1973.

Duża liczba wyników teoretycznych w lingwistyce ogólnej i formalnej otrzymana przez uczonych radzieckich w początkowym okresie badań nad tłumaczeniem maszynowym może wynikać z prozaicznej przyczyny braków sprzętowych. Z powodu niedostatecznej liczby komputerów poprawność algorytmów tłumaczących była bardzo często sprawdzana „ręcznie”. Brano po prostu człowieka nieznanego języka docelowego i on wykonywał poszczególne kroki algorytmu (zob. Harper 1960). Wielu autorów przyznaje, że mimo tych ograniczeń osiągnięcia radzieckich uczonych w tamtym okresie badań z całą pewnością nie ustępowały osiągnięciom naukowców w innych krajach (w tym również Amerykanów), a być może były nawet wartościowsze (zob. np. Roberts, Zarechnak 1974: 2851).

#### 2.4. Metody statystyczne w językoznawstwie

Wielu badaczy radzieckich było świadomych znaczenia metod statystycznych w lingwistyce. Nikołaj D. Andriejew zastosował metody statystyczne oraz teorię mnogości

do opisu struktury języka (Андреев 1959). Sebastian K. Szaumian podkreślił wagę metod statystycznych w badaniach nad językiem, a konkretnie nad wypowiedziami (Шаумян 1958).

Naukowcom w tamtym czasie przyświecała następująca idea. Jedyne, co jest dostępne badaniom, to *parole*. Jeśli więc zbierze się odpowiednio dużo odpowiednio wybranych próbek wypowiedzi, będzie można powiedzieć coś o *langue*.

W dniach 15–21 maja 1958 r. w Moskwie odbyła się pierwsza wszechzwiązkowa konferencja na temat tłumaczenia maszynowego (*Всесоюзная конференция по машинному переводу*). Dobruszyn wygłosił referaty, w których proponował zastosowanie metod teorii mnogości do badań systemu gramatycznego języka oraz użycie teorii informacji i metod statystycznych do badania struktury języka (Добрушин 1958a, b; Рapp 1966: 63).

W 1961 r. powstał pierwszy w ZSRR podręcznik lingwistyki matematycznej autorstwa Olgi S. Achmanowej, I.A. Mielczuka, Jeleny W. Paduczewej i Rebeki M. Frumkiny (Ахманова et al. 1961). Książka została przetłumaczona na angielski i wydana w USA (Akhmanova et al. 1963). Rozdział poświęcony lingwistyce statystycznej napisała Frumkina. Zastosowania metod statystycznych w językoznawstwie były bardzo różnorodne, począwszy od najbardziej naturalnych, dotyczących rozkładów słów oraz wszelkiego rodzaju konstrukcji gramatycznych w języku, do bardziej wyszukanych, dotyczących problemów z pogranicza literaturoznawstwa. Frumkina badała rozkłady określonych słów i klas słów w tekstach (Фрумкина 1962), natomiast Władimir A. Nikonow badał statystyczny rozkład przypadków w języku rosyjskim (Никонов 1959, 1961).

Nikonow był nie tylko językoznawcą, ale także literaturoznawcą. Swoje zainteresowania połączył w pracy o rytmice w poezji Majakowskiego, którą badał przy użyciu metod statystycznych (Никонов 1958). Tematykę prozodii (metrum) w twórczości Majakowskiego podjął także Andriej N. Kołmogorow zarówno w pracach samodzielnym, jak i ze współautorami (Колмогоров, Кондратов 1962; Колмогоров 1963; Колмогоров, Прохоров 1963; Колмогоров 1965).

Metody matematyczne i statystyczne były stosowane też w lingwistyce historycznej, między innymi do sformalizowania takich pojęć jak *rodzina języków*, *język przodek* itd. przez wprowadzenie odpowiednich relacji (w sensie matematycznym) na zbiorze, którego elementami są języki danej rodziny. Statystyczne modele językowe wykorzystywane były między innymi do datowania czasu rozpadu jedności prajęzyka rozpatrywanej rodziny języków. Konstruowano modele opisujące zmiany w czasie w słowniku danego języka (Арапов, Херц 1974).

### 3. Metodologia lingwistyki matematycznej

Tłumaczenie maszynowe w latach 50. i 60. XX w. było zdominowane przez dwa główne podejścia. Pierwsze z nich to podejście empiryczne – metoda prób i błędów, często z zastosowaniem metod statystycznych, konstruowanie algorytmów *ad hoc* dla specyficznych problemów. Celem tego podejścia było jak najszybsze zbudowanie działającego systemu tłumaczącego. Z drugiej strony równolegle obserwujemy podejście teoretyczne wykorzystujące fundamentalne badania językoznawcze. Tu chodziło o skonstruowanie modeli języka, które pozwoliłyby tłumaczyć dowolne teksty bez konieczności nakładania na nie dodatkowych ograniczeń. Pamiętajmy, że to drugie teoretyczne podejście zostało poniekąd wymuszone brakami sprzętowymi.

Lata 70. XX w. okazują się czasem konfrontacji dwóch naukowych koncepcji. W pierwszej, którą można nazwać inżynierską, dominowało praktyczne podejście do maszynowego tłumaczenia. Czołowym przedstawicielem koncepcji inżynierskiej był R.G. Piotrowski. Druga była podejściem teoretycznym, propagowanym między innymi przez I.A. Mielczuka i Ju.D. Apresjana. Uważali oni, że głównym zadaniem lingwisty jest podanie dokładnego opisu języka, który może stać się podstawą algorytmu tłumaczącego, a nie ulepszanie niedoskonałych systemów tłumaczenia maszynowego (Pevzner 2001: 86).

W latach 80. XX w. pojawiły się komputery osobiste. To spowodowało zwrot w dotychczasowych kierunkach badań w lingwistyce matematycznej. Stosunkowo łatwy dostęp do komputerów przyczynił się do pojawienia się lingwistyki korpusowej.

W latach 90. dokonał się prawdziwy przewrót w lingwistyce komputerowej (Hirst 2013: 716), który objął wszystkie kraje, gdzie prowadzone są badania nad lingwistyką komputerową. Zaczęły dominować badania ilościowe zorientowane na duże zbiory danych. Lingwistyka korpusowa wydzieliła się z lingwistyki komputerowej i rozwija w sposób od niej niezależny. Lingwistyka komputerowa zajmuje się teraz głównie problemami sztucznej inteligencji.

### 4. Po upadku imperium

Po rozpadzie ZSRR lingwistyka matematyczna jest uprawiana w Rosji i krajach byłego Związku Radzieckiego, lecz nie może się równać z czasami świetności w epoce socjalistycznej. Aby to zauważyć, wystarczy pobieżnie zapoznać się z artykułami publikowanymi w najważniejszych czasopismach poświęconych lingwistyce matematycznej, np. „Computational Linguistics” czy „Journal of Quantitative Linguistics”. Można się przekonać, jak mało jest publikacji autorstwa Rosjan, których główną afiliacją są instytucje naukowe w Rosji.

Przyczyn takiego stanu rzeczy jest kilka. Wśród nich z pewnością niskie nakłady na naukę, a co za tym idzie emigracja naukowców z byłych krajów ZSRR na Zachód.



## Literatura

- AKHMANOVA O.S., MEL'CHUK I.A., FRUMKINA R.M., PADUCHEVA E.V., 1963, *Exact Methods in Linguistics Research*, Berkeley – Los Angeles.
- ALPAC, 1966, *Language and machines: computers in translation and linguistics. A report by the Automatic Language Processing Advisory Committee*, Washington.
- ANDREEV N., 1967, *The intermediary language as the focal point of machine translation*, [w:] A.D. Booth (red.), *Machine Translation*, Amsterdam, s. 1–29.
- APRESJAN JU.D., BOGUSLAVSKIJ I.M., IOMDIN L.L., LAZURSKIJ A.V., SANNIKOV V.Z., TSINMAN L.L., 1992, *ETAP – 2: the linguistics of a machine translation system*, „Meta” nr (1) 37, s. 97–112.
- BAR-HILLEL Y., 1960, *The present status of automatic translation of languages*, „Advances in Computers” 1, s. 45–76.
- BAUDOIN DE COURTENAY J.N., 1909, *Zarys historii językoznawstwa, czyli lingwistyki (glottologii)*, Warszawa.
- BAUDOIN DE COURTENAY J.N., 1927/1984, *Ilościowość w myśleniu językowym*, [w:] idem, *O języku polskim*, „Biblioteka Filologii Polskiej”, Warszawa, s. 423–443.
- BELSKAYA I.K., 1957, *Machine translation of languages*, „Research” 10, s. 383–389.
- GRZYBEK P., 2004, *History of quantitative linguistics. Nikolaj Gavrilovič Černyševskij. A Forerunner of Quantitative Stylistics in Russia*, „Glottometrica” 7, s. 91–93.
- GRZYBEK P., KELIH E., 2004, *Anton Semënovič Budilovič (1846–1908). A Forerunner of Quantitative Linguistics in Russia*, „Glottometrica” 7, s. 94–96.
- GRZYBEK P., KELIH E., 2005, *Zur Vorgeschichte quantitativer Ansätze in der russischen Sprach- und Literaturwissenschaft*, [w:] R. Köhler, G. Altmann, R.G. Piotrovski (red.), *Quantitative Linguistics. An International Handbook*, Berlin, s. 23–64.
- HARPER K., 1960, *Soviet research in machine translation*, [w:] H. P. Edmundson (red.), *Proceedings of the National Symposium on Machine Translation Held at the University of California, Los Angeles, February 2–5, 1960*, Englewood Cliffs, s. 1–12.
- HIRST G., 2013, *Computational linguistics*, [w:] K. Allan (red.), *The Oxford Handbook of the History of Linguistics*, Oxford, s. 707–726.
- HUTCHINS J. (red.), 2000, *Early Years in Machine Translation: Memoirs and biographies of pioneers*, „Studies in the History of the Language Sciences” 97, Amsterdam – Philadelphia.
- HUTCHINS J., 2006, *Machine Translation: History*, [w:] K. Brown (red.), *Encyclopedia of Language & Linguistics*, t. 7, Oxford, s. 375–383.
- HUTCHINS J., 2010, *Machine translation: a concise history*, „Journal of Translation Studies” nr (1–2) 13, s. 29–70.
- HUTCHINS J., 2013, *Historical sketch of machine translation in Eastern and Central Europe*; [on-line:] <http://www.hutchinsweb.me.uk/Hamburg-2013.pdf>.
- KELIH E., GRZYBEK P., 2005, *Neuanfang und Etablierung quantitativer Verfahren in der sowjetischen Sprach- und Literaturwissenschaft (1956–1962)*, [w:] R. Köhler, G. Altmann, R.G. Piotrovski (red.), *Quantitative Linguistics. An International Handbook*, Berlin, s. 65–82.
- KEMPGEN S., 1995, *Russische Sprachstatistik. Systematischer Überblick und Bibliographie*, München.
- KOLMOGOROFF A., 1933, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin.



- МОЛОՏՆԱՅԱ Т.Н., 2000, *My memoirs of MT in the Soviet Union*, [w:] J. Hutchins (red.), *Early Years in Machine Translation: Memoirs and biographies of pioneers*, „Studies in the History of the language Sciences” 97, Amsterdam – Philadelphia, s. 227–232.
- PAPP F., 1966, *Mathematical linguistics in Soviet Union*, The Hague.
- ПЕВЗНЕР В., 2001, *First steps of language engineering in the USSR: The 50s through 70s*, „International Journal of Translation” nr (1–2) 13, s. 81–86.
- ПИОТРОВСКИ R.G., 1984, *Text-Computer-Mensch*, Bohum.
- ПИОТРОВСКИ R.G., 2000, *Machine translation in the former Soviet Union and in the Newly Independent States (NIS). Prehistoric, romantic era, prosaic time*, [w:] J. Hutchins (red.), *Early Years in Machine Translation: Memoirs and biographies of pioneers*, „Studies in the History of the language Sciences” 97, Amsterdam – Philadelphia, s. 233–242.
- ПИОТРОВСКИ R., РОМАНОВ Y., 1999, *Machine translation in the former Soviet Union and in the Newly Independent States (NIS)*, „Histoire Épistémologie Langage” nr (1) 21, s. 105–117.
- РЕВЗИН I.I., 1966, *Models of language*, London.
- РОБЕРТС А.Н., ЗАРЕЧНАК М., 1974, *Mechanical translation*, [w:] T.A. Sebeok, A.S. Abramson, D. Hymes, H. Rubenstein, E. Stankiewicz (red.), *Current Trends in Linguistics*, t. 12, The Hague – Paris, s. 2825–2870.
- СЕДЛАЧІКОВА В., 2012, *Historie matematické lingvistiky*, Brno.
- СОКОЛОВА S., 1997, *STYLUS – Outline of Functions and Features*, [w:] *MT Summit VI. Machine Translation: Past, Present, Future. Proceedings, 29 October – 1 November 1997*, San Diego – California, s. 257–258; [on-line:] <http://www.mt-archive.info/MTS-1997-Sokolova.pdf>.
- УЭЙ А., 1990, *Mel'čuk's explanatory-combinatorial dictionary and its relevance for machine translation*, [w:] J. McNaught (red.), *Proceedings of a Workshop on Machine Translation, July 1990*, Manchester.
- Аверкиев Д.В., 1893, *О драме*, Санкт-Петербург.
- Андреев Н.Д., 1959, *Моделирование языка на базе его статистической и теоретико-множественной структуры*, [w:] *Тезисы Конференции по математической лингвистике (Ленинград 1959)*, Ленинград, s. 15–22.
- Арапов М.В., Херц М.М., 1974, *Математические методы в исторической лингвистике*, Москва.
- Ахманова О.С., Мельчук И.А., Падучева Е.В., Фрумкина Р.М., 1961, *О точных методах исследования языка*, Москва.
- Бельская И.К., Королев Л.Н., Панов Д.Ю., 1959, *Переводная машина П.П. Троянского: сборник материалов о переводной машине для перевода с одного языка на другие, предложенной П.П. Троянским в 1933 г.*, Москва.
- Будилович А.С., 1883, *Начертание церковнославянской грамматики, применительно к общей теории русского и других родственных языков*, Варшава.
- Буныковский В.Я., 1847, *О возможности введения определённых мер доверия к результатам некоторых наук и преимущественно статистике*, „Современник” III, отд. II, s. 36–49.
- Гладкий А.В., 1973, *Математические методы изучения естественных языков*, „Труды Математического института им. В.А. Стеклова” 133, s. 95–108.
- Гладкий А.В., 2005, *Размышления о взаимодействии лингвистики и математики*, Доклад на Второй международной конференции по модели «Смысл ↔ Текст»; 23–25 июня 2005, Москва, [on-line:] <http://elementy.ru/lib/164549?context=20615>.

- Добрушин Р.Л., 1957, *Элементарная грамматическая категория*, „Бюллетень Объединения по проблемам машинного перевода” 5, s. 19–21.
- Добрушин Р.Л., 1958a, *Значение математических методов в лингвистике*, [w:] *Тезисы Конференции по машинному переводу*, Москва, s. 14–15.
- Добрушин Р.Л., 1958b, *Опыт определения понятия грамматической категории*, *Тезисы Конференции по машинному переводу*, Москва, s. 37.
- Жолковский А.К., Леонтьева Н.Н., Мартемьянов Ю.С., 1960, *О принципиальном использовании смысла при машинном переводе*, [w:] *Питання прикладної лінгвістики. Тези доповідей Міжвуз наук. Наукової конф. 22–28 вересня 1960 року*, Чернівці, s. 97–99.
- Жолковский А.К., Леонтьева Н.Н., Мартемьянов Ю.С., 1961, *О принципиальном использовании смысла при машинном переводе*, [w:] *Машинный перевод. Сборник трудов ИТМ и ВТ*, Москва, s. 17–46.
- Журавлёв Ю.И., 1980, А.А. Ляпунов и становление кибернетики в нашей стране, [w:] А.А. Ляпунов, *Проблемы теоретической и прикладной кибернетики*, Москва, s. 4–17.
- Зализняк А.А., 1964, *К вопросу о грамматических категориях рода и одушевленности в современном русском языке*, „Вопросы языкознания” nr (4) 13, s. 25–40.
- Колмогоров А.Н., 1936, *Основные понятия теории вероятностей*, Москва – Ленинград.
- Колмогоров А.Н., 1963, *К изучению ритмики Маяковского*, „Вопросы языкознания” nr (4) 12, s. 64–71.
- Колмогоров А.Н., 1965, *Замечания по поводу анализа ритма «Стихов о советском папоротнике» Маяковского*, „Вопросы языкознания” nr (3) 14, s. 70–76.
- Колмогоров А.Н., Кондратов А.М., 1962, *Ритмика поэм Маяковского*, „Вопросы языкознания” nr (3) 11, s. 62–74.
- Колмогоров А.Н., Прохоров А.В., 1963, *О дольнике современной русской поэзии*, „Вопросы языкознания” nr (6) 12, s. 84–96.
- Кравец Л.Г., Василевский А.Л., Дубитская А.М., 1967, *Экспериментальная система автоматического перевода публикаций из американского патентного еженедельника „Official Gazette”*, „Научно-техническая информация. Серия 2” 1, s. 35–40.
- Кулагина О.С., 1958a, *Машинный перевод с французского языка*, „Известия высших учебных заведений. Математика” nr (6) 5, s. 46–51.
- Кулагина О.С., 1958b, *Об одном способе определения грамматических понятий на базе теории множеств*, „Проблемы кибернетики” 1, s. 203–214.
- Кулагина О.С., 1958с, *Об одном способе определения грамматических понятий*, [w:] *Тезисы конференции по машинному переводу*, s. 50.
- Кулагина О.С., 1977, *О роли А.А. Ляпунова в развитии работ по машинному переводу в СССР*, „Проблемы кибернетики” 32, s. 59–71.
- Кулагина О.С., Мельчук И.А., 1956, *Машинный перевод с французского языка на русский*, „Вопросы языкознания” nr (5) 5, s. 111–121.
- Ляпунов А.А., Кулагина О.С., 1955, *Использование вычислительных машин для перевода с одного языка на другой*, „Природа” 8, s. 83–85.
- Марков А.А., 1913, *Пример статистического исследования над текстом «Евгения Онегина»*, иллюстрирующий связь испытаний в цепь, „Известия Императорской академии наук. Серия 6” nr (3) 7, s. 153–162.

- Мельчук И.А., 1958а, *О машинном переводе с венгерского языка на русский*, „Проблемы кибернетики” 1, s. 222–264.
- Мельчук И.А., 1958b, *Некоторые выводы общего характера в связи с машинным переводом с венгерского языка*, „Бюллетень Объединения по проблемам машинного перевода” 6, s. 34–43.
- Мельчук И.А., 1959, *Статистика и зависимость рода французских существительных от их окончания*, [w:] *Вопросы статистики речи*, Ленинград, s. 112–130.
- Мельчук И.А., 1962, *Об алгоритме синтаксического анализа языковых текстов (Общие принципы и некоторые итоги)*, „Машинный перевод и прикладная лингвистика” 7, s. 45–87.
- Мельчук И.А., 1974, *Опыт теории лингвистических моделей «Смысл-Текст»*, Москва.
- Мельчук И.А., Жолковский А.К., 1984, *Толково-комбинаторный словарь современного русского языка. Опыты семантико-синтаксического описания русской лексики*, Вена.
- Никонов В.А., 1958, *Ритмика Маяковского*, „Вопросы литературы” 7, s. 89–108.
- Никонов В.А., 1959, *Статистика надежей русского языка*, „Машинный перевод и прикладная лингвистика” 3, s. 45–65.
- Никонов В.А., 1961, *Борьба надежей*, „International Journal of Slavic Linguistics and Poetics” IV, s. 13–34.
- Панов Д.Ю., 1954, *Перевод с одного языка на другой при помощи машины: Отчёт о первом успешном испытании*, „Реферативной журнал «Математика»” 10, s. 75–76.
- Панов Д.Ю., 1956а, *Автоматический перевод*, Москва.
- Панов Д. Ю., 1956b, [= D. Panov] *On the Problem of Mechanical Translation*, „Mechanical Translation” nr (2) 3, s. 42–43.
- Панов Д.Ю., 1956c, [= D. Panov] *Concerning the problem of machine translation of languages*, Moscow.
- Пиотровский Р.Г., 1975, *Текст, машина, человек*, Ленинград.
- Ревзин И.И., 1958а, „Активная” и „пассивная” грамматика Л.В. Щербы и проблемы машинного перевода, [w:] *Тезисы Конференции по машинному переводу*, Москва, s. 23–25.
- Ревзин И.И., 1958b, *Формальная теория предложения*, [w:] *Тезисы конференции по машинному переводу*, Москва, s. 50–52.
- Ревзин И.И., 1959, *О понятии «множества отмеченных фраз» в теоретико-множественной концепции О.С. Кулагиной*, [w:] *Тезисы Собрания по математической лингвистике*, Ленинград, s. 27–28.
- Ревзин И.И., 1960, *О некоторых понятиях так называемой теоретико-множественной концепции языка*, „Вопросы языкознания” 6, s. 88–94.
- Ревзин И.И., 1962, *Модели языка*, Москва.
- Ревзин И.И., Розенцвейг В.Ю., 1964, *Основы общего и машинного перевода: учебное пособие для студентов институтов и факультетов иностранных языков*, Москва.
- Томашевский Б.В., 1929, *Пятистопный ямб Пушкина*, [w:] *idem, О стихе. Статьи*, Ленинград, s. 138–253.
- Успенский В.А., 1957, *К определению части в теоретическо-множественной системе языка*, „Бюллетень Объединения по проблемам машинного перевода” 5, s. 22–26.
- Успенский В.А., 1998, *Серебряный век структурной, прикладной и математической лингвистики в СССР и В.Ю. Розенцвейг: Как это начиналось (заметки очевидца)*,

[w:] Д.А. Поспелов, Я.И. Фет (red.), *Очерки истории информатики в России*, Новосибирск, s. 273–309.

Фрумкина Р.М., 1962, *О законах распределения слов и классов слов*, [w:] Т.Н. Молошная (red.), *Структурно-типологические исследования*, Москва, s. 124–133.

Чернышевский Н.Г., 1855, *Сочинения Пушкина*, „Современник” nr 2, 3, 7, 8.

Шаумян С.К., 1958, *Структурная лингвистика как имманентная теория языка*, Москва.

### **A history of mathematical linguistics in Russia and in the USSR, with a particular regard to machine translation**

#### **Summary**

Our goal is to follow and analyse the development of mathematical linguistics in Russia and in the USSR. The first application of mathematical methods, primarily statistical, occurred in Russia already in mid-19th century; however a true bloom of mathematical linguistics happens in mid-20th century, owing to the initiation of intensive research into machine translation. Both the rapid development of machine translation in the initial phase, and later curtailment of research due to considerable financial limitations, has a political-historical justification. We finish our description of the history of linguistic research with the use of mathematical methods with the moment of the collapse of the USSR. Additionally, we show the evolution of the methodology of scientific research within mathematical linguistics.