Stanisław Szpakowicz

Autoreferat

1. Wykształcenie – posiadane dyplomy

06.1965 **42 Liceum Ogólnokształcące im. Marii Konopnickiej** w Warszawie
Świadectwo maturalne.

06.1970 **Uniwersytet Warszawski**, Wydział Matematyki, Informatyki i Mechaniki*
Dyplom ukończenia studiów magisterskich w zakresie **matematyki**, specjalność **matematyka stosowana**, z wynikiem bardzo dobrym.
Praca magisterska o automatycznej odmianie czasowników polskich została napisana pod kierunkiem prof. dr. hab. Stanisława Waligórskiego.

27.09.1978 **Uniwersytet Warszawski**, Wydział Matematyki, Informatyki i Mechaniki
Dyplom uzyskania stopnia doktora nauk matematycznych uzyskany na podstawie rozprawy doktorskiej *Automatyczna analiza składniowa polskich zdań pisanych* napisanej pod kierunkiem prof. dr. hab. Stanisława Waligórskiego.

2. Zatrudnienie

— **Wydział Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego**
  — 10.1970–12.1978 stanowisko asystenta w wymiarze pełnego etatu,
  — 12.1978–02.1983 stanowisko adiunkta w wymiarze pełnego etatu,
  — 02.1983–06.1984 stanowisko starszego programisty w wymiarze pełnego etatu,

— **Wydział Informatyki Uniwersytetu Ottawskiego, Kanada**: 06.1984–08.1984 stanowisko współpracownika naukowego (ang. *research associate*) w wymiarze pełnego etatu,

— **Wydział Informatyki Uniwersytetu Kentucky, USA** 09.1984–04.1985 stanowisko adiunkta wizytującego (ang. *visiting assistant professor*) w wymiarze pełnego etatu,

— **Wydział Informatyki Uniwersytetu Ottawskiego, Kanada**
  — 07.1985–04.1994 stanowisko profesora nadzwyczajnego (ang. *associate professor*) w wymiarze pełnego etatu,
  — 05.1994–04.1997 stanowisko profesora zwyczajnego (ang. *professor*) w wymiarze pełnego etatu;

— **Wydział Informatyki Uniwersytetu Witwatersrand, Południowa Afryka**: 07.1990–06.1991 stanowisko profesora wizytującego (ang. *visiting professor*) w ramach rocznego urlopu naukowego z Uniwersytetu Ottawskiego,

— **Szkoła Biznesu Uniwersytetu Carleton, Kanada**: 07.1994–06.2000 stanowisko profesora pomocniczego (ang. *adjunct professor*),

* W roku 1965, Wydział Matematyki i Fizyki; od roku 1968, Wydział Matematyki i Mechaniki; od roku 1975, Wydział Matematyki, Informatyki i Mechaniki

— **Instytut Kognitywistyki Uniwersytetu Carleton, Kanada** 07.1996–06.2017 stanowisko profesora pomocniczego (ang. *adjunct research professor*),

— **Szkoła Technologii i Inżynierii Informacji Uniwersytetu Ottawskiego**: 05.1997–09.2011 stanowisko profesora zwyczajnego (ang. *professor*) w wymiarze pełnego etatu,

— **Wydział Informatyki Uniwersytetu Waikato, Nowa Zelandia**: 07.1998–06.1999 stanowisko profesora wizytującego (ang. *visiting professor*) w ramach rocznego urlopu naukowego z Uniwersytetu Ottawskiego,

— **Szkoła Elektroniki i Informatyki Uniwersytetu Ottawskiego** 10.2011– stanowisko profesora zwyczajnego (ang. *professor*) w wymiarze pełnego etatu,

— **Instytut Podstaw Informatyki PAN**: 04.2005– stanowisko profesora zagranicznego w wymiarze 1/4 etatu,

3. Podstawowe osiągnięcie

Jako podstawowe osiągnięcie pt. *Język negocjacji elektronicznych* przedkładam następujące prace:

— M. Shah, M. Sokolova, S. Szpakowicz (2006). Process-Specific Information for Learning E-negotiation Outcomes, *Fundamenta Informaticae* **74**(2-3), 351-373.
— M. Sokolova, M. Shah, S. Szpakowicz (2006). Comparative Analysis of Text Data in Successful Face-to-Face and Electronic Negotiations. *Group Decision and Negotiation* **15**(2), 127-140.
— V. Nastase, S. Köszegi, S. Szpakowicz (2007). Content Analysis Through the Machine Learning Mill. *Group Decision and Negotiation* **16**(4), 335-346.
— M. Sokolova, S. Szpakowicz (2007). Strategies and Language Trends in Learning Success and Failure of Negotiation, *Group Decision and Negotiation* **16**(5), 469-484.

— M. Sokolova, V. Nastase, S. Szpakowicz, M. Shah (2005). Analysis and models of language in electronic negotiations. M. Dramiński, P. Grzegorzewski, K. Trojanowski, S. Zadrożny (red.) *Issues in intelligent systems. Models and Techniques.* Akademicka Oficyna Wydawnicza EXIT, Warszawa 2005, 197-211.

— M. Shah, M. Sokolova, S. Szpakowicz (2004). The Role of Domain-Specific Knowledge in Classifying the Language of E-negotiations, *Proc. ICON 2004, International Conference on Natural Language Processing*, Hyderabad, Indie, 99-108.
— M. Sokolova, V. Nastase, S. Szpakowicz (2004). Language in Electronic Negotiations: Patterns in Completed and Uncompleted Negotiations, *Proc. ICON 2004, International Conference on Natural Language Processing*, Hyderabad, Indie, 142-151.
— M. Sokolova, S. Szpakowicz, V. Nastase (2004). Using Language to Determine Success in Negotiations: A Preliminary Study. *Proc. 17th Conf. of the CSCSI, AI 2004*, London, Ontario, Lecture Notes in Artificial Intelligence **3060**, Springer, Berlin/Heidelberg, 449-453.
— M. Sokolova, S. Szpakowicz (2005). Analysis and Classification of Strategies in Electronic Negotiations. *Proc. 18th Conf. of the CSCSI, AI 2005*, Victoria, BC, Lecture Notes in Artificial Intelligence **3501**, Springer, Berlin/Heidelberg, 145-157.
— M. Sokolova, V. Nastase, M. Shah, S. Szpakowicz (2005). Feature Selection for Electronic Negotiation Texts, *Proc. RANLP 2005, Conf. on Recent Advances in Natural Language Processing*, Borovets, Bułgaria, wrzesień 2005, 518-524.
— M. Sokolova, S. Szpakowicz (2006). Language Patterns in the Learning of Strategies from Negotiation Texts. *Proc. 19th Conf. of the CSCSI, AI 2006*, Québec City. Lecture Notes in Artificial Intelligence **4013**, Springer, Berlin/Heidelberg, 288-299.
— M. Sokolova, V. Nastase, S. Szpakowicz (2008). The Telling Tail: Signals of Success in Electronic Negotiation Texts. *Proc. Third International Joint Conf. on Natural Language Processing IJCNLP 2008*, Hyderabad, Indie, 257-264.

Prace te składają się na projekt poświęcony językowi negocjacji elektronicznych. Cztery pierwsze to artykuły w czasopismach z listy A MNiSW, kolejna to rozdział w książce o zasięgu międzynarodowym, pozostałych siedem to artykuły konferencyjne (trzy z nich opublikowane w serii LNAI Springera, indeksowanej przez ISI Web of Science). Listy współautorów stanowią załącznik numer 6 do wniosku o przeprowadzenie postępowania habilitacyjnego.

Uwaga: cztery prace w czasopismach, trzy w materiałach Springera i jedna konferencyjna (wszystkie, których egzemplarze w PDF udało mi się pozyskać), wraz z wybranymi kilkudziesięcioma innymi moimi publikacjami, są do pobrania na Internecie:
`www.eecs.uottawa.ca/~szpak/selected_publications_for_download/`

3.1. Uwaga wstępna i zastrzeżenie

Projekt, którego wyniki przedstawiam jako moje podstawowe osiągnięcie, zakończył się sześć lat temu. Projekt ten opierał się na zbiorze tekstów powstałych ze swobodnej wymiany krótkich tekstów o naturze podobnej do fragmentów, jakie obecnie wędrują między użytkownikami cyberprzestrzeni. Zbiór ten może się dzisiaj wydawać niczym wobec masy danych tekstowych pojawiających się na portalach społecznościowych w ciągu godziny. W okresie jednak, kiedy mój projekt dobiegał końca, Facebook był zjawiskiem lokalnym mającym niewiele ponad rok, a Twitter dopiero powstawał. Próba ekstrapolacji wyników tego projektu na dzisiejszą rzeczywistość byłaby kusząca, ograniczę się wszelako do omówienia – niejednokrotnie w czasie teraźniejszym – założeń, metod, wyników i bibliografii w takiej formie, jaką miały one w tamtym czasie. Nie chcę wyciągać wniosków, które wykraczałyby poza końcowy moment projektu.

3.2. Teksty w elektronicznych negocjacjach biznesowych

Na potrzeby niniejszego dokumentu zamierzam przyjąć intuicyjne rozumienie pojęcia negocjacji; dokładna definicja nie jest niezbędna dla prezentacji analizy językowej, właściwego tematu tego projektu. Negocjacje pomiędzy dwiema albo kilkoma stronami to wymiana stanowisk, której celem jest zazwyczaj osiągnięcie kompromisu.[1] Negocjacje elektroniczne to dowolny proces negocjacji prowadzony za pomocą mediów elektronicznych – zazwyczaj za pośrednictwem e-mailu albo odpowiednio zorganizowanego interfejsu sieciowego (Thompson & Nadler 2002). Elektroniczne negocjacje biznesowe stanowią podtyp takich negocjacji o szczególnie obiektywnym charakterze, co ułatwia ich automatyczną analizę. Rzecz jasna, mają one wiele cech negocjacji biznesowych w ogólności (Cellich & Jain 2004).

Wymianę komunikatów za pośrednictwem mediów elektronicznych cechuje niejaka bezosobowość. Brakuje im pewnych właściwości negocjacji prowadzonych twarzą w twarz, podczas telekonferencji czy za pośrednictwem telefonu, a mianowicie subtelnych wzajemnych zależności właściwych kontaktom osobistym. Z drugiej strony, brak bezpośrednich osobistych nacisków ze strony bardziej wpływowych, często dominujących uczestników negocjacji może wpływać na demokratyzację całego procesu i zwiększać asertywność jego uczestników. Zauważono subtelne przesunięcia wpływów poszczególnych uczestników negocjacji (Ströbel 2000).

Media elektroniczne są kanałem komunikacji znacznie węższym niż to, na co pozwala kontakt osobisty. Brakuje w nich pewnych aspektów komunikacji interpersonalnej (Hargie & Dickson 2004). Nakłanianie do ustępstw odbywa się nie w bezpośredniej argumentacji, tylko za pomocą tekstów przesyłanych tam i z powrotem pomiędzy negocjatorami. Czyni to wiadomości tekstowe towarzyszące negocjacjom elektronicznym gatunkiem wartym badań. Takie wiadomości służą wielorakim celom: przekonaniu strony przeciwnej do przyjęcia oferty, ustaleniu przyjaznej stopy relacji poprzez niezobowiązującą rozmowę na tematy niezwiązane z przedmiotem negocjacji, wyrażeniu gotowości dojścia do porozumienia, i tak dalej. Język wiadomości tego typu można badać po to, aby stwierdzić, jak takie cele są osiągane, w jaki sposób wyrażenia językowe funkcjonują w tej dziedzinie. Metody klasyfikacji tekstów (Sebastiani 2002), nawet te stosunkowo proste, mogą prowadzić do użytecznych spostrzeżeń. Da się na przykład przewidzieć sukces czy niepowodzenie negocjacji na dość wczesnym etapie procesu albo wykryć sygnały przesunięcia wpływów pomiędzy uczestnikami.

Specjaliści z zakresu negocjacji prowadzili ręczne badania zapisów negocjacji, na przykład stosując techniki analizy zawartości (ang. *content analysis*) (Krippendorff 1980). Mnie jednak interesuje analiza automatyczna na większą skalę, wykorzystująca metody przetwarzania języka naturalnego. Ilość danych tekstowych sprzyjająca takim pracom stała się dostępna dzięki systemowi *Inspire*.

[1] Pozwolę sobie skierować Czytelnika do Wikipedii. Pierwszy akapit w `en.wikipedia.org/wiki/Negotiation` trafnie podsumowuje typowe intuicje.

3.3. Dane z systemu *Inspire*

Projekt Negoplan – patrz punkt 4.8 – przekształcił się we wczesnych latach dwutysięcznych w wielostronną inicjatywę badawczą, znaną obecnie jako *Centrum Badawcze InterNeg* (ang. *InterNeg Research Centre*).[2] Inicjatywa ta przyciągnęła duży interdyscyplinarny zespół specjalistów z kilku krajów, którzy pracują nad różnorodnymi zagadnieniami w dziedzinie wspierania decyzji (ang. *decision support*), teorii negocjacji i praktyki negocjacyjnej. Byłem jednym z głównych wykonawców dużego projektu, finansowanego z kanadyjskich funduszy federalnych w latach 2002–2006, poświęconego negocjacjom elektronicznym, mediom i prowadzeniu interakcji socjoekonomicznych.[3] Odpowiadałem w nim za jeden z tematów projektu, „Wzorce komunikacyjne i analiza tekstu". Dało mi to możliwość przebadania unikatowego i trudnego zbioru tekstów. (Zbiór ten jest naprawdę wyjątkowy. Nie istniały żadne stosowne badania, na których można by się było oprzeć – brakowało wcześniejszych systematycznych prac dotyczących tekstów elektronicznych negocjacji biznesowych, wykorzystujących metody przetwarzania języka naturalnego.)

Inspire jest to internetowy system wspierania negocjacji,[4] należący do szerokiego wachlarza zastosowań projektu InterNeg. Został on pomyślany jako narzędzie do ćwiczenia umiejętności prowadzenia negocjacji. Jego wczesna wersja była stosowana na wielu uniwersytetach na całym świecie do szkolenia studentów biznesu. Stawia on przed wszystkimi to samo proste zadanie: kupujący i sprzedający części rowerowe mają osiągnąć kompromis; jest zaledwie kilka kwestii podlegających negocjacji, a każda z nich ma przypisaną krótką listę dopuszczalnych wartości numerycznych. Ponad 2500 par nowicjuszy w zakresie negocjacji stawiło czoła temu zadaniu, dzięki czemu udało się zebrać silnie ustrukturyzowane dane o uczestnikach wraz z dokonaną przez nich wymianą ofert, a także swobodne dane tekstowe. Te drugie to krótkie teksty angielskie bez ograniczeń formy, opcjonalnie dodawane do formalnej transmisji oferty. Teksty te zawierają w sumie ponad 1,5 miliona tokenów. Każdej negocjacji przypisany jest fakt zakończenia sukcesem bądź niepowodzeniem, dzięki czemu dysponujemy przejrzystą anotacją, która jest dużą pomocą w eksperymentach maszynowego uczenia się realizowanych na tym zbiorze danych.

3.4. Projekt analizy tekstu

Projekt poświęcony negocjacjom elektronicznym był koordynowany przez Szkołę Biznesu im. Molsona (ang. *Molson School of Business*) na Uniwersytecie Concordia w Montrealu; prace nad analizą tekstu wykonywaliśmy w Ottawie w latach 2003–2006. W skład mojego zespołu badawczego wchodzili: moja doktorantka Marina Sokolova i stażystka podoktorska Vivi Nastase, a także doktorant Mohak Shah specjalizujący się w metodach maszynowego uczenia się. Zespół współpracował też przez jakiś czas z prof. Sabine Köszegi z Politechniki Wiedeńskiej (Technische Universität Wien). Kontakt z innymi badaczami pracującymi nad całym projektem był sporadyczny, a nasza grupa działała niezależnie i publikowała samodzielnie.

Zasadniczym celem projektu było wypracowanie metodologii działania na zbiorach tekstów, których przykładem są dane *Inspire*. Dane te są efektem wielokulturowej inicjatywy nauczania negocjacji. Fakt ten zachęcił nas do wyjścia poza standardowe metody analizy korpusowej i klasyfikacji tekstów, i do rozważenia, jak można takie metody ulepszyć wykorzystując szczególne właściwości negocjacji elektronicznych. Dążyliśmy też jednak do ogólności, do opracowania metod analizy tekstów przydatnych nie tyko dla tego nieco nietypowego zbioru tekstów. Badania nasze szły w kilku kierunkach i obejmowały:

[2] `interneg.concordia.ca/`
[3] `interneg.concordia.ca/views/bodyfiles/enegotiation/`
[4] `invite.concordia.ca/inspire/`

— analizę sposobów, na jakie strategie negocjacji uzewnętrzniają się w tekstach wymienianych podczas negocjacji, opartą na semantyce leksykalnej, a w szczególności przebadanie wyborów leksykalnych w danych *Inspire*;
— przestudiowanie wykonalności przepowiadania wyniku negocjacji na podstawie statystycznie rozpoznawalnych sygnałów językowych (takie przepowiednie powinny następować znacznie przed zakończeniem negocjacji);
— przebadanie różnic i podobieństw pomiędzy elektronicznymi negocjacjami biznesowymi a porównywalnymi bezpośrednimi negocjacjami biznesowymi;
— opracowanie procedury selekcji cech na podstawie wiedzy o dziedzinie i przetestowanie tej procedury na danych *Inspire*;
— rozważenie zastosowania metod maszynowego uczenia się do częściowej automatyzacji analizy kontekstu w zakresie elektronicznych negocjacji biznesowych.

Poniżej omówię szczegółowo każdy z tych wzajemnie powiązanych podprojektów.

Skuteczność naszych zadań zależała od właściwego doboru narzędzi i zasobów. Rozważaliśmy między innymi metody statystycznego modelowania języka (Manning & Schütze 1999), metody porównywania korpusów (Kilgarriff 2001), zasoby leksykalno-semantyczne, listy słów wyrażających konieczność, intencje, żądanie i tym podobne (zebrane z różnych źródeł), algorytmy maszynowego uczenia się w ogólności (Witten & Frank 2000; Sebastiani 2002), a w szczególności techniki selekcji cech (Guyon & Elisseeff 2003; Forman 2003). Studiowaliśmy również wybraną literaturę na temat negocjacji, elektronicznych negocjacji i elektronicznych negocjacji biznesowych (Brett 2001; Cellich & Jain 2004; Hargie & Dickson 2004; Herring 2001; Kersten 2000; Kersten & Noronha 1999; Ströbel 2000; Thompson & Nadler 2002). Lingwistycznych właściwości języka jako narzędzia komunikacji szukaliśmy w pracy (Leech & Svartvik 2002).

3.5. Wyrażanie strategii w tekstach dotyczących negocjacji

Negocjatorzy wykonują określone ruchy w celu osiągnięcia swoich celów i dojścia do porozumienia. Żądają informacji, argumentują, grożą, perswadują, przypochlebiają się, i ogólnie starają się wpłynąć na stronę przeciwną; można w istocie mówić tu o strategii wywierania wpływu (Marwell & Schmitt 1967). Negocjacje elektroniczne nie różnią się istotnie od negocjacji bezpośrednich jeśli idzie o występowanie takich działań (Köszegi, Srnka & Pesendorfer 2006), odróżnia je jednak sposób realizacji tych działań. Stają się one aktami językowymi, zapisanymi w tekstach towarzyszących negocjacjom elektronicznym, i można je analizować pod kątem występowania sygnałów wywierania wpływu.

Poszukujemy zatem wyrazów, zwanych przez nas strategicznymi, które mają zasadniczy udział w formułowaniu takich aktów językowych jak na przykład akceptacja, odmowa, argumentacja, perswazja czy uzasadnienie. Wsród zidentyfikowanych przez nas wyrazów są czasowniki wolitywne, czasowniki aktywności umysłowej, czasowniki modalne, wybrane przymiotniki, wyrażenia zaprzeczenia i zaimki osobowe; dalsze szczegóły i wiele innych składników tego podprojektu przedstawia praca (Sokolova & Szpakowicz 2007). Wyrazy te – przede wszystkim czasowniki modalne i rozmaite postaci wyrażeń przeczących – grają kluczową rolę w takich aktach jak polecenie, żądanie, porada, propozycja, zakaz, zaprzeczenie i tak dalej (Leech 1983). Stworzyliśmy wzorce językowe i powiązaliśmy je ze stanowiskiem rozmówców i ich nastawieniem do procesu negocjacji. Na przykład wzorzec I can / may / will sygnalizuje zaangażowanie lub zobowiązanie, You might / could / should wyraża propozycję lub poradę o zmiennej sile, a You cannot /should not / do not komunikuje sprzeciw.

Wzorce językowe odnoszą się, rzecz jasna, nie tylko do negocjacji elektronicznych: niosą one podobną informacje w wielu postaciach komunikacji za pośrednictwem wymiany tekstów. Gdy rozmówcy są negocjatorami, możemy badać częstotliwość i rozkład występowania wzorców i w ten sposób określać, jak się rozwijają negocjacje i dokąd zmierzają. Jako materiał testowy posłużyły dane *Inspire*. Okazało się, że najczęściej spotykane postaci wzorców to *you can accept, I would be, you can see, we can make*

i *I cannot accept*. Tak więc na przykład rozpowszechnionym ruchem negocjacyjnym jest propozycja (sugestia), i jest to ruch typowy dla negocjacji biznesowych, w tym elektronicznych. Wszechobecność zaimków osobowych w częstych wzorcach wskazuje na komunikację interpersonalną, i rzeczywiście taka jest natura negocjacji elektronicznych. Zauważalna obecność takich rzeczowników jak *offer* i czasowników takich jak *accept* znów jasno wskazuje negocjacje.

Wzorce językowe umożliwiają systematyczne badanie dynamiki, historii i wyników negocjacji. Jednym z możliwych zastosowań jest przepowiadanie ostatecznego rezultatu negocjacji na podstawie rozkładu wzorców językowych w ich fazie początkowej.

3.6. Przepowiadanie sukcesu i niepowodzenia negocjacji

Negocjacje mogę się zakończyć się sukcesem albo niepowodzeniem. Przepowiedzieć wynik zawczasu to ważne zadanie, szczególnie gdy chodzi o system wspierania negocjacji taki jak *Inspire*. Jeśli dostatecznie wcześnie rozpozna się ryzyko wyniku negatywnego – braku porozumienia albo załamania się negocjacji – można zaproponować sposoby zapobieżenia niepowodzeniu. Kersten & Zhang (2003) stwierdzili, że wymiana ofert na wczesnym etapie zapowiada pozytywny efekt negocjacji, pozostawienie zaś wymiany ofert na końcówkę negocjacji zazwyczaj prowadzi do niepowodzenia. Nastase (2006) poszerzyła te wyniki, także pracując na kompletnych sekwencjach formalnie ustrukturyzowanych interakcji zebranych przez system *Inspire*. To, czy przepowiadanie wyniku negocjacji jest możliwe na podstawie samych nieformalnych tekstów towarzyszących negocjacjom pozostaje ważnym pytaniem badawczym.

Problem jest trudny do rozwiązania w ogólności, bo nie bardzo wiadomo, na czym oprzeć analizę bez nadzoru (ang. *unsupervised*) tekstów typowych dla negocjacji. Obecność sygnałów językowych nie zapewnia zakończenia negocjacji sukcesem. Mogą się one załamać pomimo pozytywnych perspektyw albo po prostu może zabraknąć czasu na ich dokończenie. Czystym trafem, dane tekstowe systemu *Inspire* opatrzone są jednoznacznymi etykietami: negocjacje mogą być doprowadzone do końca (zakończone sukcesem) lub przerwane (zakończone niepowodzeniem). Tak więc pozyskaliśmy wyjątkową możliwość przeprowadzenia maszynowego uczenia się pod nadzorem.

Jest wiele wartych rozważenia cech języka i reprezentujących je rozkładów wzorców. Przetestowaliśmy kilka intuicyjnych hipotez (Sokolova, Szpakowicz & Nastase 2004). Występowanie wyrazów powiązanych z negocjowaniem – które stanowią reprezentację tekstów negocjacji właściwą dla takich procesów (ang. *process-specific*) (Shah, Sokolova & Szpakowicz 2006) – jest wyraźniejsze w negocjacjach zakończonych sukcesem. Fazy wstępne negocjacji zakończonych sukcesem i niepowodzeniem przebiegają odmiennie. Wyrażenia grzecznościowe występują częściej w negocjacjach zakończonych sukcesem. Zastosowaliśmy kilka standardowych metod maszynowego uczenia się, w tym maszyny wektorów wspierających (ang. *SVM*), drzewa decyzyjne, także drzewa jednopoziomowe (ang. *Decision Stumps*), których implementacje dostępne są na platformie Weka (Witten & Frank 2001), oraz maszynę list decyzyjnych (ang. *Decision List Machine*) (Sokolova, Marchand, Japkowicz & Shawe-Taylor 2003). Osiągnęliśmy blisko 10% poprawy trafności (ang. *accuracy*) wobec punktu odniesienia (ang. *baseline*) (wszystkie negocjacje zakończone sukcesem) równego 60%.

Inny zestaw eksperymentów (Sokolova & Szpakowicz 2007) miał za cel porównanie zachowania kilku metod reprezentacji tekstów pod kątem klasyfikacji negocjacji na zakończone sukcesem bądź niepowodzeniem: kiedy przepowiadanie wyniku negocjacji na podstawie ich fazy wstępnej jest statystycznie zbliżone do przepowiadania na podstawie całego tekstu? Punkt odniesienia – na poziomie reprezentacji i realizacji – stanowią wyrazy o najwyższej frekwencji w kolekcji tekstów. Jedna z reprezentacji, oparta na wiedzy, wykorzystuje wzorce językowe omówione powyżej, inna – występujące w tekście wyrażenia warunkowe. Zastosowaliśmy trzy standardowe metody maszynowego uczenia się: maszyny wektorów wspierających, drzewa decyzyjne i naiwny klasyfikator Bayesa, do reprezentacji tekstów złożonych z pierwszej połowy negocjacji i do całych negocjacji. Eksperymenty pokazały, że reprezentacje oparte

na wiedzy dają wiarygodne przepowiednie, bardziej stabilne, gdy stosują się do pierwszej połowy niż do całego tekstu, i że są one nieznacznie lepsze od reprezentacji stanowiącej punkt odniesienia.

Transkrypcje pierwszej połowy negocjacji dla danych *Inspire* dają jedynie nieznaczną poprawę w przepowiadaniu ich wyniku. Z drugiej strony, w tekstach stanowiących zapis negocjacji twarzą w twarz, wzorce językowe z pierwszej połowy negocjacji są bardziej przydatne do przepowiadania wyniku od wzorców występujących w drugiej połowie (Simons 1993). Musi zatem istnieć wyraźna różnica w przebiegu negocjacji elektronicznych i twarzą w twarz. Istotnie, Sokolova, Nastase & Szpakowicz (2008) znaleźli wyraźną różnicę: wzorce językowe w późniejszej części tekstów negocjacji elektronicznych wyraźnie wskazują ostateczny wynik negocjacji. Przeprowadziliśmy klasyfikację tekstów fragmentów negocjacji różnej długości (jak poprzednio, za pomocą maszyn wektorów wspierających, drzew decyzyjnych i naiwnego klasyfikatora Bayesa). Teksty były reprezentowane za pomocą wyrazów powiązanych z tematyką negocjacji oraz wyrazów strategicznych. Gdy rozważaliśmy krótsze końcowe fragmenty transkrypcji, wyrazy powiązane z tematyką negocjacji wyraźniej wskazywały wynik negocjacji, a przepowiednie cechowała wyższa trafność niż w przypadku dłuższych fragmentów z początkowej fazy procesu.

Bardziej szczegółowa analiza wyników wykazała poprawę działania wszystkich klasyfikatorów na najkrótszych końcowych fragmentach negocjacji zakończonych niepowodzeniem. Innymi słowy, w klasie tej trendy stają się bardziej wyraźne wraz ze zbliżaniem się momentu zakończenia negocjacji. Z drugiej strony, wszystkie klasyfikatory wykazują pogorszenie wyników klasyfikacji negocjacji zakończonych sukcesem, gdy końcowe fragmenty negocjacji stają się coraz krótsze. Oznacza to, że negocjacje zakończone sukcesem stają się coraz bardziej zróżnicowane wraz ze zbliżaniem się do końca, przez co wykrycie trendów staje się trudniejsze.

3.7. Negocjacje biznesowe twarzą w twarz i elektroniczne

Wiele jest istotnych i wiele powierzchownych różnic między negocjacjami bezpośrednimi a elektronicznymi, ale jest też wiele pouczających cech wspólnych. Aby odkryć, jak takie podobieństwa objawiają się w języku, porównaliśmy teksty negocjacji elektronicznych typu *Inspire* i teksty powiązane z negocjacjami o podobnym charakterze (Sokolova, Shah & Szpakowicz 2006). W tym celu wybraliśmy zestaw danych *Cartoon* (Brett 1998, 2001). Jest to zbiór transkrypcji nagrań zakończonych sukcesem negocjacji twarzą w twarz (przeprowadzanych synchronicznie podczas jednego spotkania) pomiędzy kupującym i sprzedającym program telewizyjny. W 20 takich negocjacjach uczestniczyli Japończycy, a w 20 Amerykanie. Dla porównania, spośród danych *Inspire* wybraliśmy teksty towarzyszące negocjacjom zakończonym sukcesem (przeprowadzanym asynchronicznie w okresie do trzech tygodni). Te dwie kolekcje są z grubsza podobne, tyle że dane *Cartoon* są znacznie mniejsze.

Porównanie przeprowadzono etapami. Zaczęliśmy od rozpatrzenia wyrazów rzadkich, występujących raz lub dwa w całej kolekcji. Obliczyliśmy kilka miar zróżnicowania słownictwa: współczynnik typ-okaz (ang. *the type-token ratio*), stopę wzrostu (ang. *the growth rate*) i współczynnik Sichela (ang. *Sichel's characteristic*, Sichel 1986). Policzyliśmy ponadto udział bigramów i trigramów. Liczby te, wraz z dodatkowym porównaniem z trzema innymi korpusami, nasunęły nam wniosek, że teksty bezpośrednich i elektronicznych negocjacji pomiędzy sprzedającym a kupującym są zbliżone pod względem słownictwa, a podobieństwo to jest istotne statystycznie. Z drugiej strony, obie te wyspecjalizowane kolekcje różnią się od zbiorów tekstów ogólnych, zawierają bowiem więcej rzeczowników, głównie terminów specyficznych dla negocjacji. Szczegóły można znaleźć w pracy (Sokolova, Shah & Szpakowicz 2006).

Powyższe eksperymenty w analizie korpusowej pokazały, jak ważne jest, że dane *Cartoon* w transkrypcjach negocjacji japońskich i amerykańskich różnią się rozmiarem. Stworzyliśmy więc próbki o jednolitym rozkładzie, dwie wyselekcjonowane z japońskich i amerykańskich danych *Cartoon*, jedną z całego zbioru negocjacji *Inspire* i jedną spośród negocjacji *Inspire* zawierających dłuższe teksty.

Wskaźniki dotyczące słownictwa znów pokazały, że dane *Cartoon* i *Inspire* są podobne. Następnie dla każdego z czterech zbiorów danych zbudowaliśmy model trigramowy Knesera-Ney'a oraz model trigramowy Gooda-Turinga z wygładzeniem Katza (Chen & Goodman 1996). Model Knesera-Ney'a systematycznie przewyższał model Gooda-Turinga. Entropia krzyżowa (ang. *cross-entropy*) pokazała, że, z punktu widzenia przewidywalności wyników, japońskie dane *Cartoon* są bliższe danym *Inspire* niż dane amerykańskie.

Zbudowaliśmy też model krzyżowy dwóch zbiorów *Cartoon* i dwóch zbiorów *Inspire*. Przewidywanie wyników dla jednego zbioru *Inspire* wytrenowanego na drugim przebiegało podobnie, trenowanie natomiast na japońskich danych *Cartoon* dawało lepsze przewidywania dla danych amerykańskich niż odwrotnie. Wyniki krzyżowego modelowania danych *Cartoon* na podstawie *Inspire* i *vice versa* nie były jednoznaczne. Na koniec przebadaliśmy informację specyficzną dla procesu we wszystkich czterech zbiorach danych.[5] Jak można było się spodziewać po przebiegu wcześniejszych eksperymentów, dwa zbiory danych *Cartoon* różnią się względem bigramów charakterystycznych zawierających wyrazy specyficzne dla procesu, zaś zbiory danych *Inspire* są pod tym względem podobne.

3.8. Dobór cech właściwych dla procesu negocjacji

Negocjacje można traktować jako szczególnego rodzaju proces komunikacji interpersonalnej: jako ciąg kroków wymiany stanowisk. Gdy ma się do czynienia wyłącznie z tekstami odpowiadającym takim stanowiskom i dąży się do automatycznego uczenia się z tych tekstów, przydaje się skuteczna reprezentacja tekstów. Wadą najbardziej chyba znanej ogólnej metody reprezentacji tekstów, tzw. *worka słów* (ang. *Bag-of-Words*), jest wysoka wielowymiarowość, niewygodna, gdy danych jest bardzo mało. Procedura doboru cech (ang. *feature selection*) zmniejsza, często dość drastycznie, liczbę wymiarów, usuwając cechy, które mają mały udział w podnoszeniu jakości uczenia się (dla tekstów dotyczących negocjacji, uczenie się często oznacza klasyfikacje albo kategoryzację tekstów).

Nie zamierzam przeładowywać niniejszego dokumentu szczegółami dotyczącymi podstaw. Guyon & Elisseeff (2003) i Forman (2003) dokonali przeglądu tematyki doboru cech; na tym przeglądzie się opieraliśmy.[6] Przegląd metod kategoryzacji tekstów podał Sebastiani (2002), a podstawy metod kategoryzacji doboru cech (ang. *categorisation of feature selection methods*), wliczając w to ocenę wybranych cech, opisali Liu & Yu (2005).

Shah, Sokolova & Szpakowicz (2006) wprowadzili procedurę doboru dla danych tekstowych cech właściwych dla danego procesu (ang. *process-specific*). Została ona opracowana dla komunikacji między dwiema osobami, stosuje się też jednak w innych sytuacjach. Taka komunikacja za pośrednictwem komputera (ang. *computer-mediated communication*) (Herring 2001) ma pewne przydatne właściwości. Procedura buduje kilka modeli N-gramowych, a następnie koryguje je kładąc nacisk nacisku na te właściwości. W szczególności, teksty mogą być nieformalne czy wręcz niedbałe. Ponadto, w przeciwieństwie do tekstów w ogólności, takich jak korpus Browna (Francis & Kučera 1967),[7] wśród wyrazów o wysokiej frekwencji mogą się znaleźć wyrazy znaczące (ang. *content words*) (rzeczowniki, czasowniki, przymiotniki, przysłówki) – jak zauważyliśmy w punkcie 3.7; wprowadzona przez nas procedura doboru skupia się na rzeczownikach. Kluczowym zasobem dla tej procedury jest korpus angielszczyzny ogólnej, według którego konstruuje się modele N-gramowe.

Procedura wyszukuje wyrazy znaczące, które są częste w danych dotyczących komunikacji i rzadkie w angielszczyźnie ogólnej. Rozkład unigramowy dostarcza punktu odcięcia (ang. *cut-off point*) dla wysokich frekwencji. Rzeczowniki znalezione wsród wyrazów o najwyższej frekwencji przyjmuje się za zalążki (ang. *seeds*) dla dalszych kroków procedury. Buduje ona modele N-gramowe (bigramowe i

[5] Zastosowaliśmy metodę omówioną w punkcie 3.8.

[6] Artykuły pochodzą z bardzo użytecznego wydania specjalnego *Journal of Machine Learning Research* poświęconego doborowi zmiennych i cech ⟨`jmlr.csail.mit.edu/papers/special/feature03.html`⟩.

[7] Zobacz także `icame.uib.no/brown/bcm.html` – znajduje się tam dokumentacja *on-line*.

trigramowe), które pomagają w poszukiwaniu cech za pomocą techniki bootstrapu. Cechy-kandydaci muszą występować dostatecznie często blisko zalążków, to znaczy w N-gramach o wysokiej frekwencji, które zawierają zalążki. Aby nie wybierać niestosownych cech, procedura odrzuca te cechy-kandydatów, dla których N-gramy o najwyższej frekwencji zawierają wyrazy funkcyjne. Ostatnia faza procedury zapewnia, że ostatecznie dobrane cechy to te, których najczęstsze N-gramy zawierają rzeczowniki zalążkowe. Formalny opis procedury i szczegółowe omówienie eksperymentów mających na celu potwierdzenie jakości doboru cech właściwych dla danego procesu można znaleźć w pracy (Shah, Sokolova & Szpakowicz 2006).

Procedura doboru cech właściwych dla danego procesu jest całkowicie ogólna. Działa ona dla dowolnych dwóch korpusów, z których jeden (zazwyczaj mniejszy) z założenia zawiera język standardowy, zaś drugi – język niestandardowy. Procedura jest niezależna od języka. Zastosowaliśmy ją, jak można by się spodziewać, do danych *Inspire* (i korpusu Browna), uzyskując 123 wyrazy właściwe dla negocjacji. W ten sposób każdą negocjację można reprezentować za pomocą wektora 124 częstości, po jednej dla każdej cechy i dodatkowej dla zliczenia wszystkich pozostałych tokenów. Reprezentację tę porównaliśmy z kilkoma innymi w zadaniu klasyfikacji negocjacji na zakończone sukcesem albo niepowodzeniem. Skonstruowaliśmy trzy zbiory cech. Pierwszy zbiór opiera się na algorytmie przeszukiwania best-first, drugi – na przyroście informacji (ang. *information gain*), a trzeci składa się z wyrazów specyficznych dla nieformalnej rozmowy, które występują jedynie w tekstach *Inspire* niepowiązanych z kwestią negocjacyjną. Zastosowaliśmy cztery typowe klasyfikatory: naiwny klasyfikator Bayesa, maszynę wektorów wspierających, drzewa decyzyjne i maszynę list decyzyjnych. Spośród wyników szeroko zakrojonych eksperymentów, wspomnę tylko o jednym: cechy właściwe dla procesu działają lepiej od pozostałych trzech zestawów cech dla wszystkich czterech klasyfikatorów, chociaż różnica jest często nieznaczna.

Pozwolę sobie zakończyć ten punkt krótką uwagą: cechy właściwe dla procesu mogą być punktem wyjścia do interesujących analiz semantycznych. Podjęliśmy próbę przypisania takim cechom kategorii semantycznych. Poszukiwaliśmy cech zgodnych z ideami zawartymi w pracy (Hargie & Dicksona 2004).[8] Działaliśmy na unigramach, wykorzystując bigramy i trigramy do niewyszukanego ujednoznaczniania (ang. *word-sense disambiguation*) unigramów. Automatyczne rozpoznawanie kategorii unigramów było wspomagane przez wersję *on-line* słownika *Longman Dictionary of Contemporary English* (Summers 2003). Uzyskane znaczniki semantyczne zostały ręcznie zaopatrzone w odnośniki do naszych kategorii docelowych. Efektem końcowym tej pracy był niewielki słownik semantyczny danych *Inspire*.

3.9. Metody maszynowego uczenia się dla analizy zawartości

Przez cały czas trwania projektu, mój zespół badawczy miał tylko sporadyczny kontakt z innymi grupami uczestniczącymi w projekcie-matce, który był poświęcony negocjacji elektronicznych, mediów i prowadzenia interakcji socjoekonomicznych. W roku 2005 zorganizowaliśmy spotkanie szerszej grupy: workshop poświęcony *analizie formalnej i nieformalnej wymiany informacji podczas negocjacji*.[9] Współprzewodniczyłem temu workshopowi i byłem gościnnie współredaktorem dwóch poworkshopowych numerów specjalnych czasopism.

Pewna współpraca doszła jednak do skutku, a jeden niewielki podprojekt doprowadził do powstania publikowalnych wyników (Nastase, Köszegi & Szpakowicz 2007). Była to działalność uboczna z ciekawą przesłanką. Rozpatrywaliśmy zastosowanie technik maszynowego uczenia się w sytuacji, która zdawała się wymagać głębokiego przetwarzania ręcznego, na dodatek w środowisku nieobeznanym z takimi metodami.

[8] Znaleźliśmy w danych dziesięć kategorii: powiązane z negocjacjami, badania, zwroty nieformalne, właściwe dla *Inspire*, zainteresowania, nazwy osobowe, adresy emailowe, adresy miejsc, wyrazy funkcyjne, inne.

[9] `nebel.site.uottawa.ca/workshop/workshop.html`

Srnka & Köszegi (2007) pokazują, jak negocjacje elektroniczne transkrybuje się na materiał tekstowy, dzieli na jednostki, kategoryzuje i koduje . Jednostką może być pojedyncza wypowiedź, zdanie, zakończona myśl albo inny fragment takiego typu. Jednostkom przypisano dziewięć kategorii.[10] Dowolny częściowo zautomatyzowany system analizy zawartości powinien w takiej sytuacji rozpoznać fragmenty tekstu, które reprezentują jednostki komunikacyjne, wykryć temat każdej jednostki oraz wydobyć wzorce komunikacyjne. Eksperymentowaliśmy z wykrywaniem tematu, wybierając drzewa decyzyjne (ze względu na małą liczbę danych) i naiwny klasyfikator Bayesa. Problem jest dziewięcioklasowy, z nietrywialnym brakiem równowagi klas i punktem odniesienia ustawionym jako procent jednostek należących do danej klasy. Uzyskaliśmy mimo to obiecujące wyniki uczenia (Nastase, Köszegi & Szpakowicz 2007): znacznie przewyższyły one punkt odniesienia. Dokładność (ang. *precision*) i pełność (ang. *recall*) były najwyższe dla kategorii protokół komunikacyjny i zachowanie *stricte* negocjacyjne. Były to najliczniejsze klasy z przejrzystymi wzorcami językowymi, które wyrażają zachowanie i nastawienie. Zachowanie taktyczne, kategoria o najniższej dokładności, jest w istocie trudno zinterpretować nawet osobom dokonującym kodowania.

3.10. Spostrzeżenia i wnioski

Mógłbym poczynić wiele pouczających uwag o badaniu języka negocjacji elektronicznych, chcę jednak tylko odnotować szczęśliwy zbieg okoliczności. Znakomite możliwości badawcze pojawiają się, gdy spotykają się właściwe teorie, właściwe narzędzia i właściwe dane. Skorzystaliśmy też wielce z wnikliwych spostrzeżeń badaczy ze środowiska, które bada komunikację międzyludzką, negocjacje i w szczególności negocjacje elektroniczne. Mieliśmy do dyspozycji obszerny zbiór narzędzi do przetwarzania języka naturalnego, zwłaszcza do analizy korpusowej, i szeroki wybór metod maszynowego uczenia się właściwych dla danych językowych. Co najważniejsze, mieliśmy dostęp do intrygującego a zarazem ambitnego zbioru danych.

[10] Zachowanie *stricte* negocjacyjne, zachowanie nastawione na zadania, argumentacja perswazyjna, zachowanie taktyczne, zachowanie afektywne, komunikacja prywatna, komunikacja proceduralna, jednostki komunikacji właściwe dla danego tekstu, protokół komunikacyjny (Srnka & Köszegi 2007).

4. Pozostałe osiągnięcia

Gdybym miał wskazać motyw przewodni swojej różnorodnej kariery naukowej, wybrałbym *język*. Zawsze pochłaniała mnie lingwistyka informatyczna ze zdrową przymieszką lingwistyki formalnej, ale zajmowałem się także językami reprezentacji i językami programowania. Moja praca dydaktyczna i promotorstwo także konsekwentnie skupiały się na językach.

Przez lata pracy naukowej brałem udział w dziesięciu z górą projektach badawczych. Niniejszy dokument zorganizowałem wedle projektów, mniej więcej w porządku chronologicznym. Podstawowa cezura to moja przeprowadzka z Polski do Kanady w połowie lat 1980. Kilka tematów stanowi pomost łączący te dwa okresy. Można też powiedzieć, że wirtualnie powróciłem do domu dzięki istotnemu zaangażowaniu w prace nad tworzeniem polskiego wordnetu o swojskiej nazwie *Słowosieć*. Mieszkam już jednak w Kanadzie tak długo, że traktuję swój znaczny ale z dawna nietknięty polski dorobek przede wszystkim jako odskocznię do późniejszego życia zawodowego. Dał mi on wszelako pewną pozycję naukową, a niektóre z moich wczesnych prac są wciąż doceniane, chociaż od ich powstania upłynęły ponad trzy dziesięciolecia.

Poniżej przedstawię po kolei wszystkie swoje projekty, mniej lub bardziej dokładnie. Pominę tylko projekt poświęcony językowi negocjacji elektronicznych (chronologicznie następuje on po pracach opisanych w punkcie 4.8), zakończony kilka lat temu. Opisałem go szczegółowo w punkcie 3 jako osiągnięcie podstawowe.

Ogólne uwagi o współautorstwie

Mam niewiele publikacji, których jestem jedynym autorem. W swojej pracy naukowej z uporem przestrzegam zasady stałej współpracy. Publikowałem prace ze swoimi doktorantami i magistrantami, ze stażystami podoktorskimi pod moim współnadzorem, ze współbadaczami w projektach grupowych finansowanych przez duże granty i z partnerami we współorganizowanych ewaluacjach różnorodnych systemów analizy semantycznej.

W pracach napisanych wspólnie z moimi doktorantami i magistrantami zawsze byłem odpowiedzialny za ogólny temat i kierunek programu badań, jak też za szeroko rozumianą strukturę eksperymentów i ich ewaluacji. Wkład studentów miał najczęściej charakter techniczny, zwłaszcza w materii eksperymentów; do nich też w dużej mierze należał szczegółowy przegląd literatury. Zazwyczaj to ja odpowiadam za jakość argumentacji naukowej. Patrz punkt 3.2.2 *Działalności naukowej*, gdzie znajduje się pełna lista moich doktorantów. Jeżeli nazwisko z tej listy znajduje się pośród autorów artykułu, oznacza to taki właśnie podział wkładu we wspólną pracę.

Moje wspólne prace ze stażystami podoktorskimi, także wymienionymi w punkcie 3.2.2 *Działalności naukowej*, miały zwykle jako współautorów moich kolegów, wraz którymi nadzorowałem pracę stażysty. Odpowiedzialność intelektualna za pracę rozkładała się po równi na wszystkich autorów, a stażysta był zwykle autorem wiodącym.

W publikacjach z Maciejem Piaseckim mój współautor wnosi kluczowy wkład techniczny. Jest on niekwestionowanym liderem projektu *Słowosieć*. Wspólnie kształtujemy przekaz, a ja odpowiadam za tryb narracji. Inni współautorzy mają nietrywialny wkład w stronę techniczną i w formułowanie wniosków. W szczególności lingwiści – Magdalena Derwojedowa i Magdalena Zawisławska w pierwszym trzyleciu projektu, a Marek Maziarz i Ewa Rudnicka obecnie – odgrywali zasadniczą rolę w decyzjach czysto językoznawczych.

Wszelka inna współpraca, a zwłaszcza moja wieloletnia wpółpraca z Markiem Świdzińskim (plus niekiedy z Zygmuntem Salonim) i później z Grzegorzem Kerstenem (plus kilkoma stałym współautorami), opierała się na zasadzie równego udziału autorów. Oznacza to równy wkład intelektualny i równą odpowiedzialność za przekaz.

Prawie wszystkie moje publikacje zostały napisane po angielsku. Niemal zawsze odpowiadałem za język, styl i układ tych prac.

4.1. Analiza morfologiczna języka polskiego

Wprowadzenie w lingwistykę informatyczną zawdzięczam Januszowi Bieniowi. Nasze wczesne prace nad analizą morfologiczną języka polskiego opierały się na leksykologicznych i leksykograficznych zasadach, które sformułował Jan Tokarski (a potem podjął Zygmunt Saloni i jego zespół). To Janusz zapoznał mnie z tymi zagadnieniami. Nasza praca nie była specjalnie nowatorska w skali globalnej, ale mimo to osiągnęliśmy kilka pionierskich wyników w przetwarzaniu języka naturalnego, nie tylko dla polszczyzny. Bogata fleksja – cecha charakterystyczna języków słowiańskich – nie zajmowała wiele miejsca we wpływowych pracach nad analizą języka w USA, jako że fleksja angielska jest dość uboga. Udało się nam zaadaptować pewne wyrafinowane formalne teorie językoznawcze do przetwarzania komputerowego, i w ten sposób pokazać lingwistom, już ponad czterdzieści lat temu, jak przydatne mogą być komputery w ich dziedzinie (Łukaszewicz & Szpakowicz 1976).

Moja własna korzyść z zaangażowania w ten wczesny okres lingwistyki informatycznej w Polsce to nawiązanie bardzo dobrego roboczego kontaktu naukowego z grupą lingwistów, których nowatorskie idee szczególnie łatwo podlegały przetwarzaniu komputerowemu.

4.2. Systemy konwersacyjne

Projekt ten – prowadzony we współpracy z Januszem Bieniem i Witoldem Łukaszewiczem – to zaledwie przypis w moim dorobku publikacyjnym. Zajmowaliśmy się jednak tą tematyką przez kilka lat (i nabyliśmy doświadczenia w posługiwaniu się sprzętem komputerowym, który w krajach rozwiniętych wzbudziłby uśmiech politowania dziesięć lat wcześniej). Efektem prac nad systemem *Marysia* (Bień, Łukaszewicz & Szpakowicz 1973a, 1973b, 1974; Łukaszewicz & Szpakowicz 1974) była stosunkowo złożona struktura, przydatna w tworzeniu oprogramowania zdolnego do komunikowania się w języku polskim, z naciskiem na morfologię. Moje zainteresowanie interfejsami w języku naturalnym przetrwało projekt *Marysia* i znalazło bardziej praktyczny wyraz, kiedy zapoznałem się z językiem programowania Prolog, który jest doskonałym narzędziem do przetwarzania języka naturalnego.

4.3. Programowanie w logice i Prolog

Janusz Bień sprowadził Prolog do Polski na początku lat 1970, gdy był to jeszcze zupełnie nowy język programowania. Prologiem zajmowałem się jednak przede wszystkim wspólnie z Feliksem Kluźniakiem, z którym współpracowaliśmy blisko przez wiele lat. Pomogłem mu w budowie pięciu implementacji Prologu, coraz bardziej zaawansowanych i skutecznych. Jesteśmy współautorami dwóch książek o Prologu, pierwszej w historii (Kluźniak & Szpakowicz 1983, po polsku, która ze zrozumiałych przyczyn przeszła niezauważona poza granicami Polski) i trzeciej w historii (Kluźniak & Szpakowicz 1985, po angielsku, dobrze przyjętej, ale słabo reklamowanej) Moje zainteresowania skupiały się na metodologii programowania w logice i zastosowaniach języka Prolog (Kluźniak & Szpakowicz 1984), a także na kwestiach implementacyjnych. W pionierskich czasach Prologu napisałem kilka stosunkowo sporych programów. Najważniejszym z nich był analizator składniowy języka polskiego, który był częścią mojej rozprawy doktorskiej. Współtworzyłem także interfejs do bazy danych w języku naturalnym i implementację systemu zarządzania bazą danych; drugim z tych systemów posługiwałem się przez kilka lat na zajęciach z baz danych, które prowadziłem na Uniwersytecie Ottawskim w Kanadzie. Inny duży system zaprogramowany wyłącznie w Prologu to *Negoplan* – patrz punkt 4.8.

W późnych latach 1980 brałem udział – z ramienia Kanadyjskiego Komitetu Normalizacyjnego (ang. *Canadian Standards Association*) – w opracowywaniu międzynarodowego standardu Prologu, sponsorowanego przez ISO.

4.4. Lingwistyka i gramatyka formalna języka polskiego

Miałem szczęście stawiać pierwsze kroki w lingwistyce informatycznej w czasie, gdy niewielka ale silna grupa polskich lingwistów zajęła się opisem języka w sposób możliwie najbardziej formalny. Długa lista publikacji tej grupy zawiera tak znaczące pozycje jak (Saloni & Świdziński 2011). Niekwestionowanym przywódcą tej grupy był Zygmunt Saloni. Janusz Bień i ja reprezentowaliśmy informatyczną stronę długotrwałych wspólnych wysiłków badawczych. Mój wkład to zastosowanie notacji formalnej (gramatyka metamorficzna, zwana też gramatyką DCG (ang. *Definite Clause Grammar*)) do opisu wybranych zjawisk językowych polszczyzny. Zbudowałem formalny model składni polskiej, którego część zapisałem jako parser zaimplementowany w Prologu (Szpakowicz 1978, 1983).[11] Po upływie trzech dziesięcioleci model ten wyraźnie się zestarzał, wciąż jednak uważa się go za przydatne wprowadzenie do powierzchniowo-syntaktycznej analizy języka polskiego; (Bańko 1990) to jedna z wcześniejszych pozycji wyrażających taką opinię, a (Bień 2009, punkty 2.3-2.4 wraz z bibliografią) to dyskusja bardziej szczegółowa. Długie lata ścisłej współpracy łączyły mnie z Markiem Świdzińskim (Szpakowicz & Świdziński 1981, 1990). Zstosował on mój formalizm w swojej późniejszej pracy nad gramatyką polską (Swidzinski 1992). Opublikowaliśmy także kilka artykułów z zakresu leksykografii (Saloni, Szpakowicz & Świdziński 1982, 1990; Świdziński & Szpakowicz 1993, 1994).

4.5. Analiza składniowa

Analizator syntaktyczny, który był zaimplementowaną częścią gramatyki formalnej języka polskiego z mojej rozprawy doktorskiej, służył jedynie do weryfikacji proponowanych rozwiązań i nie był pomyślany jako gotowe narzędzie programistyczne. Dostępny w tamtych latach sprzęt komputerowy był bardzo wolny, a interpreter Prologu, jakim wówczas dysponowałem, był jeszcze wolniejszy.[12]

Moje zainteresowanie automatyczną analizą składniową przetrwało lata. Obejmuje ono wczesne próby poradzenia sobie ze składnią języków o swobodnym szyku wyrazów (Bień, Laus-Mączyńska & Szpakowicz 1980; Bień & Szpakowicz 1982); powróciłem do tego tematu znacznie później (Sayeed & Szpakowicz 2004). Opublikowałem także prace o analizie składniowej z szerokim pokryciem językowym angielskich tekstów technicznych (Delisle & Szpakowicz 1991, 1995; Copeck, Delisle & Szpakowicz 1992; Delisle, Scarlett & Szpakowicz 2000).

4.6. Pozyskiwanie wiedzy z tekstu

Mój pierwszy długotrwały projekt badawczy po przeprowadzce do Kanady dotyczył pozyskiwania wiedzy z tekstów angielskojęzycznych, przede wszystkim tekstów o charakterze objaśniającym i normatywnym. Rozpoczął on wieloletnią (trwającą półtorej dekady) współpracę z moimi kolegami z Ottawy. Moimi partnerami naukowymi byli wówczas Stan Matwin, Franz Oppacher, Rob Holte i Doug Skuce. Uzyskałem fundusze od rządu kanadyjskiego na swoje własne badania, a także na kilka dużych projektów wspólnie z wymienionymi kolegami. W projektach byli także zatrudnieni dwaj stażyści podoktorscy. Nasze eksperymentalne systemy działały zarówno w oparciu o techniki regułowe, jaki techniki wykorzystujące maszynowe uczenie się. Uzyskane wyniki zostały opisane w dwóch artykułach w czasopismach (Szpakowicz 1990; Matwin & Szpakowicz 1992) oraz w licznych artykułach konferencyjnych (Skuce, Matwin, Tauzovich, Szpakowicz & Oppacher 1985; Szpakowicz, Matwin & Skuce 1986; Tauzovich, Matwin, Oppacher, Skuce & Szpakowicz 1986; Constant, Matwin & Szpakowicz 1987; Szpakowicz 1988; Szpakowicz & Koperczak 1990; Yang & Szpakowicz 1991a, 1991b, 1994; Delannoy, Feng, Matwin & Szpakowicz 1993; Matwin & Szpakowicz 1993; Delisle, Barker, Delannoy, Matwin

[11] Wersję elektroniczną pracy (Szpakowicz 1983) można znaleźć na ⟨`ftp://ftp.mimuw.edu.pl/pub/users/polszczyzna/Szpakowicz/foszp.pdf`⟩.

[12] Gramatyka ta oraz znacznie bardziej obszerna gramatyka opisana w (Świdziński 1992) doczekały się później znacznie bardziej efektywnych implementacji: pierwsza w pracy (Wachowski 2000), druga w pracy (Woliński 2004).

& Szpakowicz 1994; Feng, Copeck, Szpakowicz & Matwin 1994; Delisle & Szpakowicz 1997; Barker, Delisle & Szpakowicz 1998).

4.7. Relacje semantyczne

Przez wiele lat ważne miejsce pośród moich zainteresowań badawczych zajmowały relacje semantyczne, i zajmują je do dzisiaj. Troje moich doktorantów zajmowało się tym fascynującym tematem. Wraz z nimi opublikowałem sporo artykułów, z których kilka było wielokrotnie cytowanych, zwłaszcza (Barker & Szpakowicz 1998) i (Nastase & Szpakowicz 2003a). Współorganizowałem kilka wspólnych zadań (ang. *shared tasks*) podczas najważniejszych konferencji poświęconych ewaluacji semantycznej, SemEval, w 2007 i 2010 roku. Współorganizuję też takie zadanie na SemEval 2013. Książka o relacjach semantycznych pomiędzy rzeczownikami, zamówiona przez wydawnictwo Morgan & Claypool, ukaże się w roku 2013; moimi współautorami są Vivi Nastase, Preslav Nakov and Diarmuid Ó Séaghdha. Relacje leksykalno-semantyczne znajdują się także w centrum moich prac nad zasobami językowymi – patrz punkt 4.11.

Poza dwiema pracami zacytowanymi powyżej, opublikowałem na temat relacji semantycznych cztery artykuły w czasopismach (Delisle, Barker, Copeck & Szpakowicz 1996; Barker, Copeck, Delisle & Szpakowicz 1997; Girju, Nakov, Nastase, Szpakowicz, Turney & Yuret 2009; Nastase & Szpakowicz 2009) i szereg artykułów konferencyjnych (Copeck, Delisle & Szpakowicz 1992; Delisle, Copeck, Szpakowicz & Barker 1993; Barker & Szpakowicz 1995; Delisle & Szpakowicz 1997; Barker, Delisle & Szpakowicz 1998; Nastase & Szpakowicz 2001a, 2005, 2006b; Nastase, Sayyad-Shirabad, Sokolova & Szpakowicz 2006; Kennedy & Szpakowicz 2007; Girju, Nakov, Nastase, Szpakowicz, Turney & Yuret 2007; Butnariu, Kim, Nakov, Ó Séaghdha, Szpakowicz & Veale 2009, 2010; Hendrickx, Kim, Kozareva, Nakov, Ó Séaghdha, Padó, Pennacchiotti, Romano & Szpakowicz 2009, 2010). Ostatnie pięć prac przedstawia mój udział w organizowaniu ewaluacji semantycznej.

Współredaguję także gościnnie numer specjalny pisma *Journal of Natural Language Engineering* poświęcony komputerowemu podejściu do semantyki rzeczowników złożonych. Numer ten ukaże się w roku 2013.

4.8. Modelowanie i wspieranie negocjacji

W latach 1986–1998 pracowałem nad podejmowaniem decyzji, analizą negocjacji i wspieraniem negocjacji (ang. *decision making, negotiation analysis, negotiation support*). Moim głównym partnerem naukowym był kierownik zespołu, Gregory Kersten. W pierwszym okresie w badaniach uczestniczyli także Wojtek Michałowski, Stan Matwin i Zbig Koperczak, w drugim okresie w projekcie brało udział trzech stażystów podoktorskich. Była to bardzo owocna współpraca specjalistów z zakresu teorii decyzji i wspierania negocjacji z ekspertami z dziedziny sztucznej inteligencji. W wyniku tych badań opracowaliśmy system analizy decyzji i symulacji podejmowania decyzji, zwany Negoplan, oparty na zasadach modelowania restrukturyzowalnego (ang. *restructurable modelling*). Negoplan – wraz z teorią, na której został oparty i założeniami reprezentacji wiedzy – można było zastosować do modelowania negocjacji, do indywidualnego i wspólnego podejmowania decyzji i do symulacji symbolicznej. System przechowywał reprezentację celów, struktury i zachowania trzech stron współuczestniczących w negocjacjach: agenta, którego decyzje są wspierane przez system, pozostałych uczestników i środowiska, w którym decyzje są podejmowane. Negoplan miał szereg praktycznych zastosowań, w szczególności w nauczaniu umiejętności stawiania diagnoz medycznych.

Summa summarum, mój udział w projekcie Negoplan zaowocował – nie wliczając następujących po nim badań nad językiem negocjacji, omówionych w punkcie 3 – szeregiem recenzowanych artykułów w czasopismach (Kersten, Michalowski, Matwin & Szpakowicz 1988; Matwin, Szpakowicz, Koperczak, Kersten & Michalowski 1989; Kersten & Szpakowicz 1990, 1994, 1995, 1997; Kersten, Szpakowicz & Koperczak 1990; Kersten, Michalowski, Szpakowicz & Koperczak 1991; Michalowski, Kersten, Ko-

perczak & Szpakowicz 1991; Koperczak, Matwin & Szpakowicz 1992) i recenzowanych artykułów konferencyjnych (Kersten, Matwin, Michalowski & Szpakowicz 1987; Kersten, Michalowski, Matwin & Szpakowicz 1987; Matwin, Szpakowicz, Kersten, Michalowski & Koperczak 1987; Szpakowicz, Matwin, Kersten & Michalowski 1987; Michalowski, Kersten, Koperczak, Matwin & Szpakowicz 1988; Matwin, Szpakowicz & Koperczak 1989; Szpakowicz, Kersten & Koperczak 1989, 1990; Koperczak, Kersten & Szpakowicz 1990; Szpakowicz & Koperczak 1990; Kersten & Szpakowicz 1991, 1992, 1993, 1994, 1998; Koperczak, Szpakowicz & Matwin 1991; Kersten, Michalowski & Szpakowicz 1992; Szpakowicz, Koperczak & Kersten 1992; Kersten, MacDonald, Rubin & Szpakowicz 1993; Szpakowicz & Kersten 1993; Kersten, Lu & Szpakowicz 1994; Kersten, Rubin & Szpakowicz 1994; Kersten, Cray & Szpakowicz 1995; Noronha & Szpakowicz 1996a, 1996b).

4.9. Streszczanie i segmentacja tekstu

Automatycznym streszczaniem tekstów interesuję się od roku 1996. Poczynając od roku 2001, mój zespół na Uniwersytecie Ottawskim regularnie uczestniczy we wspólnej ewaluacji systemów streszczających, a mianowicie Document Understanding Conference, przemianowanej kilka lat temu na Text Analysis Conference. To bardzo dla środowiska ważne coroczne wydarzenie organizowane jest przez *National Institute of Standards and Technology* (NIST).[13] W latach 2001–2011 streszczanie było jednym z zadań (ang. *track*), a kolejne zadanie planuje się na rok 2013. Przez owe ponad 10 lat moimi współpracownikami byli: koleżanki z Uniwersytetu Ottawskiego Nathalie Japkowicz i Diana Inkpen, magistranci Loïs Rigouste, Darren Kipp, Anna Kazantseva i Alistair Kennedy, a także mój wieloletni pomocnik Terry Copeck.

Poza raportami prezentowanymi rokrocznie na sponsorowanych przez NIST spotkaniach DUC/TAC, wynikiem moich badań nad streszczaniem tekstu jest artykuł w czasopiśmie (Kazantseva & Szpakowicz 2010), jedna z najlepszych prac w moim *résumé*. Opiera się on na pracy magisterskiej mojej magistrantki o streszczaniu nowel. Nasze obecne badania, składowa jej programu doktorskiego, dotyczą streszczania długich narracji, przede wszystkim powieści. W przeciwieństwie do tekstów technicznych, powieści zazwyczaj nie zawierają wyraźnych sygnałów struktury. Najbardziej obiecującą metodą znajdowania jakiejś struktury (warunek *sine qua non* skutecznego streszczania) jest segmentacja tematyczna tekstu. Opublikowaliśmy dwie prace o segmentacji na najważniejszych konferencjach w przetwarzaniu języka naturalnego, organizowanych przez Association for Computational Linguistics (Kazantseva & Szpakowicz 2011, 2012).

Mam ponadto szereg artykułów konferencyjnych o streszczaniu wiadomości prasowych (Delannoy, Barker, Copeck, Laplante, Matwin & Szpakowicz 1998; Chali, Matwin, Szpakowicz 1999; Copeck, Japkowicz & Szpakowicz 2002; Rigouste, Szpakowicz, Japkowicz & Copeck 2004; Copeck & Szpakowicz 2004; Kazantseva & Szpakowicz 2006; Kennedy & Szpakowicz 2010a, 2010b; Kennedy, Kazantseva, Inkpen & Szpakowicz 2012). Współprzewodniczyłem także dwóm workshopom poświęconym streszczaniu tekstów.

4.10. Analiza emocji

Analiza wydźwięku (ang. *sentiment analysis*) jest stosunkowo nową "modą" w przetwarzaniu języka naturalnego. Emocje, najbardziej intrygująca odmiana wydźwięku, budzą żywe zainteresowanie od połowy lat 2000; mam parę godnych uwagi wyników z okresu, kiedy formowała się ta tematyka (Aman & Szpakowicz 2007, 2008). Pierwsza z tych publikacji ma bardzo dobre cytowania, a zbiór danych stworzony w ramach tych badań jest ceniony wśród osób zajmujących się analizą emocji. W mojej obecnej pracy nad analizą emocji uczestniczą Diana Inkpen i nasza wspólna doktorantka Diman Ghazi. Wspólnie opublikowaliśmy kilka prac (Ghazi, Inkpen & Szpakowicz 2010a, 2010b, 2012), a

[13] Patrz też `duc.nist.gov` i `www.nist.gov/tac` – znajduje się tam więcej szczegółów.

teraz rozpoczynamy badania nad automatyczną analizą emocji w tekście umotywowaną kognitywnie, w szczególności nad przyczynami emocji wyrażonymi w tekście. Współredagowałem także gościnnie numer specjalny pisma *Computational Intelligence* poświęcony analizie wydźwięku.

4.11. Wordnety i inne tezaurusy

Bazy wiedzy leksykalno-semantycznej – wordnety oraz tezaurusy komputerowe – precyzyjnie reprezentują znaczenie wyrazów i relacje semantyczne między nimi. Wśród wielu innych zastosowań, ułatwiają one rozpoznanie wystąpień relacji zachodzących pomiędzy wyrazami w tekście. Takie zasoby językowe są kluczowym elementem "krajobrazu" w przetwarzaniu języka naturalnego. Zarówno badacze, jak i twórcy zastosowań komercyjnych, przywykli znajdować łatwy, bezpłatny dostęp do takich zasobów. Aby jednak zapewnić niezawodność i lingwistyczną wiarygodność, zasoby takie trzeba konstruować ręcznie, aczkolwiek ze znaczącym logistycznym wsparciem komputerowym. Poczynania tego rodzaju są od wielu lat w centrum moich zainteresowań. Od ponad 10 lat zajmuję się komputerowym tezaurusem *Roget's*, a od połowy lat 2000 – polskim wordnetem o nazwie *Słowosieć*. Drugi z tych projektów będzie jeszcze trwał dobrych kilka lat.

Różne wersje systemu *Roget's* stosowano sporadycznie do zadań z przetwarzania języka naturalnego, a kilka jego wersji (niekiedy o wątpliwym statusie co do praw autorskich) jest dostępnych on-line. Pierwszą implementację rzeczywiście nastawioną na przetwarzanie języka naturalnego, napisaną w języku Java i mającą rozbudowany interfejs programistyczny (ang. *Application Programming Interface*), stworzył mój magistrant Mario Jarmasz we wczesnych latach 2000. System został wypełniony danymi pochodzącymi z wersji *Roget's* wydawnictwa Penguin z 1987 roku. System zastosowano, z bardzo dobrymi wynikami, do kilku typowych zadań, w tym do konstrukcji łańcuchów leksykalnych (ang. *lexical chains*) i do pomiaru podobieństwa semantycznego pomiędzy wyrazami w tekście. Opublikowaliśmy kilka prac (Jarmasz & Szpakowicz 2001a, 2001b, 2003a, 2003b), a najczęściej cytowana pośród nich jest bez wątpienia (Jarmasz & Szpakowicz 2003a).

Zdecydowanie najpopularniejszą i najszerzej stosowaną bazą leksykalną dla języka angielskiego jest WordNet, zbudowany na Uniwersytecie Princeton (PWN). Mario Jarmasz i ja pokazaliśmy zalety systemu *Roget's* jako uzupełnienia dla PWN, ale jego szerokie rozpowszechnianie okazało się niemożliwe. Dane są własnością firmy Pearson Education, a właściciel nie zdecydował się na ich swobodne udostępnienie. Mój doktorant Alistair Kennedy zajął się automatycznym uaktualnieniem na stan dzisiejszy słownictwa w wersji *Roget's* z danymi z 1911 roku, które są publicznie dostępne.[14] Wyniki naszej pracy pojawiły się w kilku pracach konferencyjnych (Kennedy & Szpakowicz 2007, 2008, 2010a, 2011, 2012). System *Open Roget's*, dostępny na swobodnej licencji BSD (do ściągnięcia z `rogets.eecs.uottawa.ca`) zyskał sporą popularność.

Pomijając porównanie systemu *Roget's* z PWN w kilku typowych zadaniach, moje bezpośrednie zainteresowanie WordNetem było przez wiele lat dość ograniczone. Jestem współautorem całkiem szeroko cytowanego artykułu o ujednoznacznianiu sensu wyrazów (Li, Szpakowicz & Matwin 1995). W roku 2004 zostałem zaproszony do nowotworzonego zespołu lingwistów i lingwistów informatycznych, którego celem było stworzenie wordnetu dla języka polskiego. Główni wykonawcy pracowali na Politechnice Wrocławskiej (będącej instytucją koordynującą a zarazem miejscem pracy koordynatora Macieja Piaseckiego), na Uniwersytecie Warszawskim i w Polskiej Akademii Nauk. Kolejne projekty prowadzone były już wyłącznie na Politechnice Wrocławskiej.

Istnieją dwa rodzaje metod tworzenia wordnetu dla nowego języka. Można przetłumaczyć PWN, a następnie dokonać rozległych zmian (ang. *post-processing*), aby zniwelować różnice (często znaczne) pomiędzy dwoma językami. Można też stworzyć nowy zasób od zera. My wybraliśmy ten drugi sposób. Założenia ogólne, szczegóły techniczne – lingwistyczne i obliczeniowe – wraz z wynikami pierwszej fazy

[14] W 2006 roku Alyona Medelyan zręcznie podmieniła dane z 1987 na swobodnie dostępne dane z 1911 roku – `www.nzdl.org/ELKB` podaje dalsze szczegóły.

konstrukcji *Słowosieci* przedstawiliśmy w monografii (Piasecki, Szpakowicz & Broda 2009). Książka, dostępna także w sieci, ma spore uznanie w środowisku wordnetowym. *Słowosieć* jest dostępna na swobodnej licencji podobnej do licencji PWN. Już teraz jest drugim co do wielkości takim zasobem na świecie.

Jak dotąd, moje pozostałe recenzowane publikacje o *Słowosieci* obejmują pięć artykułów w czasopismach (Broda, Piasecki & Szpakowicz 2010; Maziarz, Piasecki, Szpakowicz & Rabiega-Wiśniewska 2011; Maziarz, Piasecki, Szpakowicz, Rabiega-Wiśniewska & Hojka 2011; Maziarz, Szpakowicz & Piasecki 2012; Maziarz, Piasecki & Szpakowicz 2013), dwa rozdziały w książkach (Piasecki, Broda, Głąbska, Marcińczuk & Szpakowicz 2009; Kurc, Piasecki & Szpakowicz 2012) i liczne artykuły konferencyjne (Derwojedowa, Piasecki, Szpakowicz & Zawisławska 2007; Piasecki, Szpakowicz & Broda 2007a, 2007b; Broda, Derwojedowa, Piasecki & Szpakowicz 2008; Broda, Piasecki & Szpakowicz 2008; Derwojedowa, Piasecki, Szpakowicz, Zawisławska & Broda 2008; Derwojedowa, Szpakowicz, Zawisławska & Piasecki 2008; Piasecki, Szpakowicz, Marcińczuk & Broda 2008; Broda, Piasecki & Szpakowicz 2009; Piasecki, Broda, Marcińczuk & Szpakowicz 2009; Kurc, Piasecki & Szpakowicz 2010a, 2010b; Piasecki, Szpakowicz & Broda 2010; Rudnicka, Maziarz, Piasecki & Szpakowicz 2012; Maziarz, Piasecki & Szpakowicz 2012a, 2012b).

Jestem gościnnie współredaktorem numeru specjalnego pisma *Language Resources and Evaluation* poświęconego wordnetom i relacjom semantycznym. Numer ten zostanie opublikowany w roku 2013.

Bibliografia (poza publikacjami wymienionymi w *wykazie publikacji*)

Mirosław Bańko (1990). Niektóre problemy oceny adekwatności gramatyk (na przykładzie fragmentu gramatyki Szpakowicza). *Studia Gramatyczne* IX, 55-72.

Janusz S. Bień (2009). *Problemy formalnego opisu składni polskiej* BEL Studio, Warszawa.

Jeanne M. Brett (1998). Inter- and Intra-cultural Negotiation: U.S. and Japanese Negotiators. *Academy of Management Journal* **5**(41), 495-510.

Jeanne M. Brett (2001). *Negotiating Globally: How to Negotiate Deals, Resolve Disputes, and Make Decisions Across Cultural Boundaries.* Jossey-Bass, San Francisco.

Claude Cellich, Subhash C. Jain (2004). *Global Business Negotiations: A Practical Guide.* Thomson, South-Western.

Stanley F. Chen, Joshua Goodman (1996). An Empirical Study of Smoothing Techniques for Language Modeling. *Proc. 34th Annual Meeting of the Association for Computational Linguistics*, 310-318.

George Forman (2003). An Extensive Empirical Study of Feature Selection Metrics for Text Classification. *Special Issue on Variable and Feature Selection, Journal of Machine Learning Research*, **3**, 1289-1305.

W. Nelson Francis, Henry Kučera (1967). *Computational Analysis of Present-Day American English.* Brown University Press.

Isabelle Guyon, André Elisseeff (2003). An Introduction to Variable and Feature Selection, Special Issue on Variable and Feature Selection, *Journal of Machine Learning Research*, **3**, 1157-1182.

Owen Hargie, David Dickson (2004). *Skilled Interpersonal Communication: Research, Theory and Practice.* Routledge.

Susan C. Herring (2001). Computer-mediated discourse. D. Tannen, D. Schiffin, H. Hamilton (red.) *The Handbook of Discourse Analysis.* Oxford, Blackwell Publishers, 612-634 (ella.slis.indiana.edu/ herring/cmd.pdf).

Gregory E. Kersten (2000). Modeling Distribution and Integrative Negotiations. Review and Revised Characterization. *Group Decision and Negotiation*, **10**(6), 493-514.

Gregory E. Kersten, Sunil Noronha (1999). WWW-based Negotiation Support: Design, Implementation, and Use. *Decision Support Systems*, **25**, 135-154.

Gregory E. Kersten, Grant Zhang (2003). Mining Inspire Data for the Determinants of Successful Internet Negotiations. *Central European Journal of Operations Research* **11**(3), 297-316.

Adam Kilgarriff (2001). Comparing corpora. *International Journal of Corpus Linguistics*, **6**(1), 97-133.

Sabine T. Köszegi, Katharina Srnka, Eva-Maria Pesendorfer (2006). Electronic Negotiations - A Comparison of Different Support Systems. *Die Betriebswirtschaft*, **66**(4), 441-463.

Klaus Krippendorff (1980). *Content Analysis. An Introduction to Its Methodology.* Sage Publications.

Geoffrey N. Leech (1983). *Principles of Pragmatics.* Longman.

Geoffrey N. Leech, Jan Svartvik (2002). *A Communicative Grammar of English.* wyd. trzecie, Longman.

Huan Liu, Lei Yu (2005). Toward Integrating Feature Selection Algorithms for Classification and Clustering. *IEEE Transactions on Knowledge and Data Engineering*, **17**(4), 491–502.

Chris Manning, Hinrich Schütze (1999). *Foundations of Statistical Natural Language Processing.* MIT Press. Cambridge, MA.

Gerald Marwell, David R. Schmitt (1967). Dimensions of Compliance-Gaining Behavior: An Empirical Analysis. *Sociometry*, **30**(4), 350-364.

Vivi Nastase (2006). Concession Curve Analysis for Inspire Data. *Group Decision and Negotiation* **15**(2), 185-193.

Zygmunt Saloni, Marek Świdziński (2011). *Składnia współczesnego języka polskiego*, wyd. 5-te, PWN (wyd. pierwsze w 1985).

Fabrizio Sebastiani (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys* **34**(1), 1-47.

Herbert S. Sichel (1986). Word Frequency Distributions and Type-Token Characteristics. *Mathematical Scientist*, **11**, 45-72.

Tony Simons (1993). Speech Patterns and the Concept of Utility in Cognitive Maps: the Case of Integrative Bargaining. *Academy of Management Journal*, **38**(1), 139-156.

Marina Sokolova, Mario Marchand, Nathalie Japkowicz, John Shawe-Taylor (2003). The Decision List Machine. *Advances in Neural Information Processing Systems*, **15**, 921-928, The MIT Press.

Katharina J. Srnka, Sabine T. Köszegi (2007). From Words to Numbers - How to Transform Rich Qualitative Data into Meaningful Quantitative Results: Guidelines and Exemplary Study. *Schmalenbach's Business Review, Die Zeitschrift für betriebswirtschaftliche Forschung* t. 59, 29-57.

Michael Ströbel (2000). Effects of Electronic Markets on Negotiation Processes. *Proc. 8th European Conference on Information Systems*, Wiedeń, t. 1, 445-452.

Della Summers, red. (2003). *Longman Dictionary of Contemporary English*, wyd. czwarte. Pearson Education.

Marek Świdziński (1992). *Gramatyka formalna języka polskiego*, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.

Leigh Thompson, Janice Nadler (2002). Negotiating via information technology: Theory and application. *Journal of Social Issues*, **58**(1): 109-124.

Adam Wachowski (2000). Adekwatność lingwistyczna analizatorów składniowych języka polskiego. Praca magisterska, Instytut Informatyki Uniwersytetu Warszawskiego.

Ian H. Witten, Eibe Frank (2000). *Data Mining: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann.

Marcin Woliński (2004). Komputerowa weryfikacja gramatyki Świdzińskiego, rozprawa doktorska, Instytut Podstaw Informatyki, Polska Akademia Nauk (`www.ipipan.waw.pl/~wolinski/publ/mw-phd.pdf`).

Stanisław Szpakowicz