

Recenzja pracy doktorskiej Konrada Gołuchowskiego

Metoda identyfikacji schematów walencyjnych w tekstach polskich

Krzysztof Jassem

11 stycznia 2019

1 Wstęp

Celem niniejszej recenzji jest określenie, czy rozprawa Konrada Gołuchowskiego zatytułowana „Metoda identyfikacji schematów walencyjnych w tekstach polskich” spełnia wymagania stawiane pracom doktorskim.

W pierwszej części recenzji poszczególne rozdziały pracy omówione są pod kątem merytorycznym oraz formalnym. W drugiej części oceniona zostaje celowość przeprowadzonych badań oraz zastosowana w nich metodologia.

2 Ocena poszczególnych rozdziałów pracy

2.1 Wprowadzenie

Wprowadzenie napisane jest bardzo poprawnie pod względem formalnym. Autor wprowadził i opatrzył klarownymi przykładami terminologię stosowaną w kolejnych rozdziałach pracy.

Autor nie uniknął jednak stwierdzeń kontrowersyjnych. Na przykład na str. 22 walencję określa się jako cechę wyłącznie czasowników, choć w wielu opracowaniach walencję przypisuje się również innym częściom mowy, np. przymiotnikom lub rzeczownikom. Uważam, że modyfikatory lepiej byłoby zdefiniować w kontraście do argumentów (poprzez stwierdzenie, że ich niewystąpienie nie powoduje niepoprawności składniowej zdania), gdyż aktualna definicja (opisywanie okoliczności takich jak czas i miejsce) stoi w sprzeczności z wieloma przykładami podanymi w pracy, w których okoliczniki czasu i miejsca są argumentami, a nie – modyfikatorami.

Ostatni akapit sekcji 1.2.1 jest chyba niepotrzebnym powtórzeniem fragmentu ze wstępu (str. 12).

Niezrozumiała dla mnie jest definicja preferencji selekcyjnych. Termin wydaje mi się określeniem typu „masło maślane” (zarówno „preferencje”, jak i „selekcja” to wyrazy bliskie znaczeniowo wyrazowi „wybór”). Bardziej adekwatne byłoby chyba określenie „preferencje semantyczne” lub „preferencje leksykalne”. Termin „preferencji selekcyjnych” definiuje się w rozprawie na trzy sposoby: 1) jako uogólnienie ograniczeń selekcyjnych, 2) jako zależności semantyczne między predykatem i jego argumentami, 3) jako zbiór wybranych pojęć z leksykonu. Ta wielość definicji może wprowadzić czytelnika w konfuzję. Pojęcie ograniczeń selekcyjnych ma obrazować Tabela 1.3, która zawiera obrazujące je przykłady. Tabela może sugerować, że przez preferencję selekcyjną należy ro-

zumić ograniczenie semantyczne narzucone na argument predykatu (argument ma należeć do danego synsetu Słownosieci). Jednak lektura Rozdziału 2. rozprawy skłania czytelnika do przyjęcia założenia, że przez preferencję selekcyjną rozumie się ograniczenie leksykalne.

W podrozdziale 1.4 autor przybliżyła postawione zadanie badawcze: automatyczny wybór schematu walencyjnego dla predykatu w zdaniu. Autor przyjmuje dwa założenia, które ułatwiają wykonanie zadania. Pierwsze z nich to pominięcie w rozważaniach zdań, w których nie są realizowane żadne argumenty. Przykład, który ma obrazować taką sytuację jest niejasny („Przyjechałem, żeby ratować Gruzję”). W zdaniu tym argument podmiotu realizowany jest przez podmiot domyślny „ja”. Nie jest więc dla czytelnika klarowne, czy w rozważaniach pomijane są tylko zdania, w których podmiot jest domyślny (tak, jak w podanym przykładzie), czy też takie, w których nie są realizowane inne argumenty poza podmiotem (np. „Jan przyjechał, żeby ratować Gruzję”),

Drugie założenie jest kontrowersyjne: „W ewaluacji pomięte są zdania, w których wybór schematu nie był jednoznaczny”. W tym miejscu recenzji należy wybiec nieco wprzód: w metodzie manualnej anotacji zastosowanej w opisywanym eksperymencie, w przypadku różnicy między dwoma anotacjami, jako wiążąca przyjmowana była decyzja superanotatora. Zgodnie z tym założeniem sytuacja niejednoznaczności nie powinna mieć w ogóle miejsca (problem ten poruszany jest w niniejszej recenzji ponownie przy omówieniu metodyki manualnej anotacji). Przyjęte założenie daje możliwość optymalizacji wyników ewaluacji poprzez arbitralną selekcję zdań uważanych za niejednoznaczne.

Jako przykład obrazujący oczekiwane działanie algorytmu Autor podaje następujące zdanie: „W Krakowie nasza Dorota kupiła sobie dwa nowe zamki”. Twierdzi on, że jako prawidłowo wybrany schemat powinien zostać wybrany taki, który oprócz podmiotu (realizowanego przez frazę „nasza Dorota”), dopełnienia bliższego w tzw. przypadku partytywnym (realizowanego przez frazę „dwa nowe zamki”) i dopełnienia dalszego („sobie”) zawiera dwa typy fraz niezrealizowanych w zdaniu: frazę przymkową z przymkiem „za” oraz (w pewnym uproszczeniu) okolicznik miejsca.

W moim odczuciu najlepszym schematem realizowanym przez to zdanie jest schemat, w którym realizowane są dokładnie trzy argumenty. W słowniku *Walenty* dostępnym na stronie internetowej projektu Clarin¹ znalazłem taki schemat dla czasownika *przynieść*, w którym jednak dopełnienie bliższe realizowane jest przez przypadek strukturalny (a nie – lepiej tu pasujący – partytywny). Nie znalazłem natomiast w tym źródle schematu podanego w pracy, lecz do niego podobny, w którym oprócz typów zrealizowanych w zdaniu zawarte są dwa różne typy okoliczników miejsca.

Przyjmując jednak, że w wersji słownika *Walenty* będącego w posiadaniu Autora, schemat walencyjny podany w rozprawie jest zawarty, pozostaje dla mnie niejasne, dlaczego schemat ten jest lepszy od schematu z trzema argumentami, czy też od tego podobnego schematu z pięcioma argumentami, który ja znalazłem.

Zdefiniowanie pojęcia „schematu najlepszego” jest kluczowe dla całej rozprawy. Tymczasem przyjęta definicja jest bardzo niejasna – anotatorzy podejmujący decyzję mieli wybrać schemat

¹<http://walenty.clarin-pl.eu/>

„pasujący”, a przypadku niejednoznaczności „schemat najczęstszy lub najbardziej intuicyjny”. Ta definicja jest za mało precyzyjna. Po pierwsze, anotator nie ma wiedzy, który schemat jest najczęstszy (ani słownik *Walenty*, ani system anotacji nie udostępniają tej informacji). Po drugie, trudno oczekiwać, aby anotator posiadał intuicję co do poprawności schematu walencyjnego, gdyż takie zadanie (w przeciwieństwie do innych, łatwiejszych wyzwań lingwistycznych) nie jest „intuicyjne”.

Moim zdaniem wybór schematu najlepszego można zdefiniować znacznie bardziej precyzyjnie, np. za pomocą następującej instrukcji:

1. Schemat z argumentem zawierającym frazę stałą (w rozprawie pojęcie to nosi nazwę: „frazę zleksykalizowaną o ustalonym brzmieniu”) ma priorytet nad schematem z argumentem z frazą zleksykalizowaną.
2. Schemat z argumentem z frazą zleksykalizowaną ma priorytet nad schematem z argumentem bez frazy zleksykalizowanej. Jeśli na przykład dane są dwa schematy walencyjne dla czasownika chodzić:

(a) NP(STR) + Lex('spać') + LEX('z kurami');

(b) NP(STR) + INFP(IMPERF) + PREPNP(Z, INS)

to dla zdania: „Jan chodził spać z kurami”, które realizuje oba schematy, pierwszeństwo powinien mieć schemat pierwszy.

3. Schemat z argumentem „częściowo zleksykalizowanym” (pojęcie wprowadzone przez autora recenzji) powinien mieć priorytet nad „schematem nieleksykalnym” (pojęcie wprowadzone przez autora recenzji). Na przykład, dla dwóch schematów **NP(STR) + XP(ADL)** oraz **NP + PREPNP(DO)** i zdania „Jan chodził do szkoły” wybrany powinien zostać schemat drugi.
4. Argument ze spełnionym ograniczeniem semantycznym ma priorytet nad argumentem bez tego ograniczenia (gdyby ograniczenia semantyczne były rzeczywiście weryfikowane).
5. W przypadku nierozstrzygniętym przez punkty 1) do 4) wybierany powinien być schemat, dla którego w zbiorze obrazujących go przykładów znajduje się zdanie, w którym predykat ma najbardziej zbliżone znaczenie do zdania analizowanego.

Powyższa instrukcja stosuje pojęcia fraz stałych i fraz zleksykalizowanych. Mam wrażenie, że w recenzowanej pracy nie określono precyzyjnie rozróżnienia tych dwóch typów fraz. Świadczyć może choćby o tym przykład podany w rozprawie, gdzie frazę „do diabła” podaje się jako przykład frazy zleksykalizowanej, a frazę „oko w oko” – jako przykład frazy stałej. Moim zdaniem obie frazy są tego samego typu – nie dopuszczają modyfikacji ani odmiany fleksyjnej żadnego ze swoich elementów składowych. Przykładem frazy zleksykalizowanej może być natomiast fraza przyimkowa „po głowie” (jako argument predykatu *chodzić*), gdyż może być ona realizowana np. w zdaniu: „Różne myśli chodziły po mojej głowie”, w którym element frazy („głowie”) jest modyfikowany przez zaimek „mojej”.

Na str. 33 popełniono prawdopodobnie błąd przy klasyfikacji cech z przytaczanego eksperymentu dla języka czeskiego. Cechy określające kategorie syntaktyczne określa się mianem cech morfologicznych, a cechy morfologiczne (dotyczące np. przypadków rzeczowników) nazywa się cechami składniowymi.

W podrozdziale 1.4.2 Autor odnosi się do podobnego eksperymentu stosującego metody nie nadzorowane. Na str. 52 i 171 postawiona jest teza, że metoda opisana w recenzowanej rozprawie wpisuje się w nurt uczenia bez nadzoru. Moim zdaniem jest to teza nieuprawomocniona. Uczenie bez nadzoru ma miejsce, gdy korpus treningowy nie jest oznaczony wynikiem klasyfikacji. W opisywanym eksperymencie korpus treningowy zawiera wyniki klasyfikacji – tyle, że są one przygotowane w sposób automatyczny.

Podsumowując, uważam, że Wprowadzenie zredagowane jest w sposób poprawny pod względem formalnym. Niektóre stwierdzenia uważam za kontrowersyjne, a opis metodologii stosowanej w pracy – za mało precyzyjny.

2.2 Rozdział 2

Rozdział 2. przedstawia autorską metodę identyfikacji schematów walencyjnych.

Na początku omawia się założenia przyjęte w eksperymencie. Nie wspomina się tutaj o założeniach omówionych we Wprowadzeniu – pominięciu zdań bez realizacji argumentów lub z realizacją niejednoznaczną. Pojawia się natomiast inne (dobrze uzasadnione) uproszczenie, polegające na rezygnacji z rozpoznawania mowy niezależnej.

Klasyczna metoda nadzorowanego uczenia maszynowego polegałaby na przygotowaniu dużych rozmiarów zbioru treningowego, w którym każdemu zdaniu przypisany jest przez człowieka dokładnie jeden schemat walencyjny. Ze względu na wysoką pracochłonność takiego rozwiązania zdecydowano się na rozwiązanie, w którym w pierwszym kroku opracowuje się pewien nieskomplikowany algorytm wyboru schematu walencyjnego (nazwijmy go algorytmem bazowym lub naiwnym). Algorytm ten wykorzystuje się następnie do automatycznego oznaczenia dużego korpusu treningowego, który staje się podstawą do wytrenowania algorytmu docelowego.

Tego typu rozwiązanie niesie za sobą następujące ryzyko: algorytm uczenia maszynowego będzie dążył do tego, aby przewidywane rezultaty były podobne do wyznaczonych przez algorytm naiwny. Eksperyment taki ma sens wtedy, gdy pomimo to rezultaty algorytmu docelowego są lepsze od algorytmu naiwnego – według pewnej obiektywnej miary.

Algorytm identyfikacji schematu walencyjnego przebiega następująco: W pierwszym kroku analizowane zdanie podlega przetwarzaniu wstępnemu, w wyniku którego zdanie podzielone jest na segmenty (odpowiadające tokenom). Każdy segment scharakteryzowany jest pod kątem morfosyntaktycznym, a ponadto rozpoznane są powierzchniowe grupy składniowe.

Następnie identyfikowane są elementy główne (syntaktyczny i semantyczny) potencjalnych podrzedników predykatów. Znalezienie takich elementów jest z punktu widzenia identyfikacji schematu walencyjnego wystarczające. Na przykład, w zdaniu „Jan przyjechał do miasta otoczonego pięć-

nyymi górami” zamiast pełnej identyfikacji dopełnienia przyimkowego („do miasta otoczonego pięknymi górami”) algorytm identyfikuje główny element syntaktyczny frazy („do”) i semantyczny („miasta”), co jest wystarczającą informacją do zidentyfikowania argumentu predykatu „przyjechać”, realizowanego w tym zdaniu. Elementy główne rozpoznawane są metodą automatyczną wytrenowaną za pomocą metody ograniczonych pól losowych z oznaczonego korpusu zdań o liczebności ponad 8 tys. elementów. Korpus ten powstał z banku drzew Składnica, w którym dla każdego zdania podane jest jego pełna reprezentacja składniowa wraz z elementami głównymi fraz. Oznaczenia fraz w Składnicy musiały zostać automatycznie skonwertowane do formatu stosowanego w słowniku *Walenty*. W trakcie tego procesu na pewno natrafiono na szereg istotnych problemów. O jednym z nich – dotyczących fraz przyimkowych – wspomniano w podrozdziale 2.1. Przykładowa fraza okolicznikowa „do Krakowa” konwertowana jest do typu PREPNP(DO, GEN) (odpowiadającego frazie przyimkowej), a nie do typu XP(ADL) (odpowiadającego okolicznikom miejsca docelowego). W tej sytuacji w korpusie trenującym typy XP są zapewne „niedoreprezentowane” (potwierdza to Rysunek 2.13), co, jak sądzę, skutkuje tym, że wytrenowany algorytm takie frazy rozpoznaje zbyt rzadko. Fakt ten może mieć niekorzystny wpływ na jakość całego procesu identyfikacji schematu walencyjnego.

Tak rozpoznane elementy główne podrzędników przypisywane są predykatom za pomocą reguł heurystycznych (np. „przypisz podrzędnik do najbliższego predykatu”). Te nieskomplikowane reguły opisane są na kilku listingach (2.1, 2.2, 2.3, 2.4), które moim zdaniem są przerostem formy nad treścią. Przy okazji tej nadmiarowej sztuczności popełnione są pewne błędy (np. niespójność nazwy dla reguły SENTPRULE, która na str.78 nazwana jest regułą dla podrzędników zdaniowych, a na str. 80 – regułą wyboru predykatu dla spójnika wprowadzającego zdanie podrzędne).

(Mam wrażenie, że podrozdział 2.2.3 został napisany w pośpiechu, bo występują w nim kolokwializmy i skróty myślowe („segmenty zmniejszając szansę na to”, „opisane dotychczas reguły wybierały”).

W następnym kroku odszukiwane są semantyczne elementy główne predykatów. Odbywa się to na podstawie ręcznie opracowanych reguł, (np. dla frazy przyimkowej wybierany jest najbliższy rzeczownik po prawej stronie przyimka występujący w przypadku wymaganym przez przyimek). Autor pisze „Umożliwia on (tenże krok, przypisek autora recenzji)...obliczenie preferencji selekcyjnych”. Nie wiem, co oznacza ta czynność (być może wynika to z mojego niezrozumienia definicji preferencji selekcyjnych). Sądziłbym, że chodzi tu o zweryfikowanie, czy semantyczny element główny podrzędnika spełnia ograniczenia semantyczne podane w schemacie walencyjnym (np. dopełnienie bliższe ma być typu „OSOBA”, co można by sprawdzić poprzez relację hiponimii w *Słownosieci*). Jednak w żadnym z algorytmów identyfikacji schematów walencyjnych, opisanych w dalszym ciągu pracy, nie znalazłem potwierdzenia tej hipotezy. Nie znalazłem również ograniczeń semantycznych w żadnym z przytaczanych w pracy schematów walencyjnych ze słownika *Walenty*. (Warto być może nadmienić, że ograniczenia semantyczne w schematach walencyjnych były rozpatrywane już w słowniku Polańskiego z roku 1985², czy w pracach dotyczących zastosowania walencji w tłuma-

²(Polański, Kazimierz (red.), 1980–1992, Słownik syntaktyczno-generatywny czasowników polskich

czeniu automatycznym³.

Wydaje mi się, że w rzeczywistości „preferencje selekcyjne” sprawdzane są wyłącznie w schematach zawierających frazy zleksykalizowane. W opisywanym eksperymencie zakłada się, że dana fraza zleksykalizowana jest zrealizowana przez podrzędnik, gdy semantyczny element główny podrzędnika jest tożsamy (lub należy do tego samego leksemu) z leksykalnym elementem frazy. Oznacza to, że preferencje selekcyjne (w rozumieniu Autora) nie są ograniczenia semantycznymi (jak wynikać to może z ich definicji), lecz ograniczeniami leksykalnymi.

Kolejnym etapem przetwarzania jest wybór schematu walencyjnego, opisany w podrozdziale 2.3. Jest to najistotniejszy fragment pracy, gdyż zawiera opis szeregu autorskich rozwiązań, począwszy od naiwnej metody bazowej, a skończywszy na kaskadzie metod zaawansowanych.

Pierwszym krokiem tego etapu jest dopasowanie oznaczeń stosowanych w *Składnicy* do oznaczeń stosowanych w słowniku *Walenty*, które wykonywane jest za pomocą algorytmu o nazwie *Match*. Nie jest dla mnie jasna współzależność pomiędzy tym algorytmem (opisanym w podrozdziale 2.3.1), a sposobem konwersji między formatami opisanym w podrozdziale 2.2.1. Wydaje mi się, że chodzi dokładnie o to samo, opisane innymi słowami. W każdym razie podrozdział 2.3.1. potwierdza moje wrażenie odniesione po lekturze podrozdziału 2.2.1, że typy fraz XP(ABL) i XP(ADL) są niedoreprezentowane, gdyż odpowiadają tylko bardzo specyficznym strukturom frazowym ze *Składnicy* (patrz Rysunek 2.2.3).

Na stronie 90. podane są przykłady argumentów zleksykalizowanych. Są one kontrowersyjne. Moim zdaniem, zarówno fraza „z kopyta”, jak i podfrazą „na oczy” w wyrażeniu „pokazywać się na oczy”, powinny być oznaczone jako frazy stałe. O tym, że niekonsekwencja w rozróżnianiu tych dwóch typów fraz może doprowadzić do błędnego przetwarzania, świadczy ostatecznie zdanie z sekcji „Argumenty zleksykalizowane” (str. 92). Zgodnie z podanym w nim algorytmem, argument zleksykalizowany „z kopyta” byłby zrealizowany przez frazę „z kopytami” w przykładowym zdaniu „Zjadłbym konia z kopytami”.

Z kolei na str. 92 jako przykład frazy stałej podano „biec na łeb, na szyję”, która w moim odczuciu powinna być zaklasyfikowana jako fraza zleksykalizowana, gdyż dopuszcza koniugację czasownika „biec”.

W podrozdziałach 2.3.2 oraz 2.3.3, 2.3.4, opisane są metody wybory schematu, które łącznie tworzą rozwiązania bazowe. Algorytm o nazwie *GreatestSchema* wybiera schemat, który pasuje do największej liczby podrzędników występujących w zdaniu. W przypadku gdy takich schematów jest więcej, algorytm nie podaje odpowiedzi rozstrzygającej, przekazując działanie kolejnej metodzie (*GreatestSchemaNoXP*), która preferuje realizację fraz przyimkowych nad frazami okolicznikowymi (np. wybór typu argumentu PREPNP(DO) nad schematem XP(ADL) dla podrzędnika „do domu”). Trzecia faza rozwiązania bazowego preferuje argumenty zleksykalizowane nad niezleksykalizowanymi.

Za pomocą kaskady tych metod (polegającej na tym, że każda kolejna z nich działa tylko na

³Jassem, K., Wagner, A. (2005), Semantic disambiguation in an MT system based on a bilingual dictionary

zdaniach nieoznaczonych przez poprzednie) automatycznie oznaczono korpus zdań, który posłużył do wytrenowania algorytmów uczenia maszynowego opisanych w kolejnych podrozdziałach.

Ostatni etap przetwarzania polegał na wyborze takiego połączenia metod uczenia maszynowego, które daje najlepsze wyniki według miary F. Zastosowano algorytm symulowanego wyżarzania, w wyniku którego uzyskano pewien zoptymalizowany ciąg algorytmów. Sekwencja ta składa się z 15-u algorytmów (str. 167 - 168), dla której w tabeli 3.26 podano odpowiednie wyniki miar dokładności, pokrycia i F-miary.

2.3 Rozdział 3

Rozdział 3. zawiera wyniki ewaluacji poszczególnych metod, a w jego skład wchodzi duża liczba tabel, zawierających wartości obliczone do czwartego miejsca po przecinku. Dla mnie wystarczające byłoby porównanie wyników rozwiązania bazowego (naiwnego) oraz docelowego, a tego porównania akurat w pracy nie ma (patrz część 4. recenzji). Taki rezultat wydałby mi się znacznie istotniejszy dla oceny wartości merytorycznej pracy niż porównanie z eksperymentami dla innych języków – szczególnie w sytuacji gdy Autor dokonał pewnych arbitralnych uproszczeń, o których była już tu mowa.

Dla mnie znacznie ciekawsze byłoby przeanalizowanie, które cechy znacząco wpływają na wyniki poszczególnych metod (dzięki temu być może udałoby się te metody ulepszyć) oraz podanie przykładów zdań i schematów, dla których poszczególne metody popełniają błędy.

3 Ocena celowości badań

Problemem badawczym opisywanym w rozprawie jest automatyczne rozpoznawanie, który spośród schematów walencyjnych czasownika podanych w słowniku *Walenty* ma miejsce w konkretnym przypadku użycia tego czasownika w zdaniu. Zazwyczaj prace doktorskie w dziedzinie nauk technicznych w specjalności informatyka rozwiązują pewien problem o konkretnym zastosowaniu praktycznym. W przypadku recenzowanej rozprawy problem badawczy wydaje mi się wyłącznie teoretyczny. Autor twierdzi, że określanie użytego schematu walencyjnego ma zastosowanie w detekcji zdarzeń i powołuje się na jedną pracę z roku 2015 „Event Extraction as Frame-Semantic Parsing”. Jednak w cytowanej pracy mowa jest o relacji między rozpoznawaniem zdarzeń a ekstrakcją semantycznych (a nie syntaktycznych) konstrukcji predykatowo-argumentowych. Ponadto Autor stawia hipotezę (za pracą z roku 1993), że czasowniki o tych samych schematach walencyjnych tworzą klasy spójne semantycznie. Teza ta jest ryzykowna. W nieco nowszej pracy⁴ próbuje się wykazać spójność semantyczną przymiotników o tej samej walencji. Praca dotyczy jednak klasy wyrazów o znacznie rzadszej skłonności do przyłączania podrzędników. Ponadto tezę starano się uzasadnić w oparciu o wspólne schematy walencyjne dla par wyrazów w dwóch językach należących do różnych rodzin.

⁴Jassem K. (2002), Semantic Classification of Adjectives on the Basis of their Syntactic Features in Polish and English

Zagadnieniem podobnym do omawianego, które w moim odczuciu miałyby realne zastosowanie, byłyby automatyczna klasyfikacja znaczeń czasowników w oparciu o schematy walencyjne. Na przykład, dla każdego z 19-u znaczeń (wg *Słowocieci*) czasownika *chodzić* określono by zestawy schematów walencyjnych (spośród 64 schematów w słowniku *Walenty*) i na podstawie automatycznej analizy podrzędników określano by prawdopodobieństwa należenia danej instancji czasownika do odpowiedniej klasy semantycznej.

Brak realnego uzasadnienia prowadzonych badań stawia pod znakiem zapytania ich celowość aplikacyjną.

4 Ocena przyjętej metodologii

Określenie, który ze schematów walencyjnych najlepiej pasuje w danym kontekście, rzadko jest oczywiste. Przykłady wyborów schematów podanych przez Autora (dla czasowników „przynieść” i „chodzić”) potwierdzają tę tezę. W ocenie automatycznego przetwarzania języka naturalnego często kierujemy się językową kompetencją człowieka. Klasyfikację uważamy za poprawną, jeśli jest zgodna z oceną mówcy danego języka. Aby na przykład zweryfikować jakość analizatora składniowego, porównujemy konstrukcje generowane przez system z anotacjami proponowanymi przez człowieka. Zakładamy przy tym, że w przypadku wieloznaczności człowiek intuicyjnie wybiera interpretację zgodną z pojmowanym przez niego znaczeniem wypowiedzi.

Zadanie wyboru schematu walencyjnego nie jest intuicyjnie pojmowane przez człowieka. Sądzę, że nawet językoznawca nie ma świadomości, że czasownik „chodzić” może (zgodnie z opisem w słowniku *Walenty*) tworzyć 64 schematy walencyjne, trudno więc spodziewać się, że intuicyjnie wybierze najlepszy z nich. Dlatego też w postawionym zadaniu należało precyzyjnie zdefiniować pojęcie najlepszego schematu walencyjnego, a tego w pracy nie uczyniono.

Przyjęto, że system uczyć się będzie na danych oznaczonych automatycznie metodą naiwną. Takie podejście jest ryzykowne, gdyż wszelkie metody uczenia dążą do upodobnienia predykcji do danych oznaczonych, a nie – do oczekiwanych. Obowiązkiem Autora metody jest wykazanie wyższości systemu wytrenowanego nad systemem naiwnym. Tymczasem w pracy, w której podano kilkanaście tabel ewaluacyjnych, zabrakło jednego porównania: F-miary rozwiązania docelowego z F-miara rozwiązania naiwnego.

Dokonałem takiego porównania samodzielnie. F-miara rozwiązania naiwnego na korpusie ewaluacyjnym (wg danych z Tabeli 3.9) wynosi 0.6082, a docelowego – 0,6620 (Str. 168). Można zatem przyjąć, że eksperyment z jednej strony wykazał wyższość metody docelowej nad metodą naiwną, a z drugiej – pozostawił niedosyt ze względu na poziom wzrostu jakości uzyskany przy znacznie wyższej złożoności rozwiązania.

5 Podsumowanie

Recenzowana rozprawa jest bardzo dobrze skomponowana. Obszerne wprowadzenie pozwala dobrze zrozumieć koncepcję pracy. Każdy rozdział poprzedzony jest wstępem oraz podsumowaniem. Język pracy jest klarowny. Nowe pojęcia wprowadzane są w logicznej kolejności – w pracy nie są stosowane terminy niezdefiniowane lub zdefiniowane dopiero po ich użyciu. Język pracy jest poprawny – w całej rozprawie znalazłem tylko kilka błędów interpunkcyjnych lub literowych. Styl nie pozostawia zastrzeżeń – zdania są zwarte i zrozumiałe.

Autor wykazał umiejętność stosowania aktualnych metod przetwarzania języka naturalnego. Potrafił przeprowadzić autorski eksperyment, unikalny na skalę międzynarodową.

Nie wskazano celowości prowadzonych badań. Nie można więc stwierdzić, do czego wykonany eksperyment może się przydać.

W metodologii popełniono zaniechanie polegające na mało precyzyjnym zdefiniowaniu pojęcia „najlepszy schemat”. Opracowane metody uczenia maszynowego dały niewysoki wzrost w stosunku do rozwiązania naiwnego.

Uważam, że walory pracy nieznacznie przeważają nad jej niedociągnięciami. Ostateczną decyzję co do akceptacji rozprawy chciałbym jednak pozostawić Radzie Instytutu po wysłuchaniu jej obrony, w której Autor mógłby się odnieść do postawianych w niniejszej recenzji zarzutów.

Krzysztof Jank