Recenzja rozprawy doktorskiej mgr Agnieszki Palion-Musioł
pt. Desambiguacion semantica automatica de los verbos seleccionados de
movimiento salir y entrar en el enfoque orientado a
objetos/Automatyczne odwieloznaczenie semantyczne wybranych
czasowników ruchu salir i entrar w ujęciu zorientowanym obiektowo

Pani mgr Agnieszka Palion-Musioł przedłożyła rozprawę pod wskazanym wyżej
tytulem wykonaną pod kierunkiem prof. dra hab. Wiesława Banysia w przewodzie
doktorskim otwartym na Wydziale Filologicznym Uniwersytetu Śląskiego. Celem
badania, jaki przed sobą postawiła jest szczegółowe wyjaśnienie mechanizmów
dezambiguacji czasowników hiszpańskich, co jest pokazane na przykładzie dwóch
wysoce polisemicznych jednostek leksykalnych oznaczających przemieszczanie: salir
oraz entrar. Wybór, jakiego dokonała jest bardzo szczęśliwy i bez wątpienia spełnia
wymogi stawiane przez odpowiednie przepisy przed dysercji doktorskimi.
Zagadnienie dezambiguacji należy do kluczowych nie tylko z punktu widzenia
teoretycznego, ale także praktycznego. W rzeczy samej, bez sprawnie funkcjonującego
rozwijania problemu odwieloznaczenia nie ma mowy o postępie w tłumaczeniu
maszynowym, bowiem program informatyczny musi zawierać wskazówkę który
odpowiednik danego wyrazu w języku źródłowym należy wybrać ze słownika
transferowego dla języka docelowego (w tym wypadku jest nim język polski). Metod
usuwania wieloznaczności jest kilka, w tym proponowana i stosowana z powodzeniem
w śląskim ośrodku naukowym metoda syntaktyczno-semanticzna wysuwająca na
pierwszy plan zorientowanie obiektowe zaproponowana przez Promotora
Doktorantki, prof. dra hab. Wiesława Banysia. Według tej metody wykonano na
Uniwersytecie Śląskim szereg prac doktorskich. Poczynając od rozpraw S. Szramek-
Karcz, B. Śmigielskiej z 2006 roku, kończąc na niedawno obronionej dysertacji M.

Nie bez kozery wskazałem strony każdej części, zabieg ten pozwala bowiem zdać sobie sprawę, że o ile sam u k l a d elementów treści należy uznać za poprawny (najpierw zagadnienia teoretyczne, potem zaś prezentacja rezultatów materiałowych), o tyle ich podział na części nie został dokonany w sposób optymalny. Na jednym poziomie zostały bowiem umieszczone części (capítulos) o znacznie zróżnicowanych rozmiarach. Obok 3-stronicowego Capítulo tercero z wnioskami, znalazł się capítulo segundo liczący blisko 400 stron! Wolalbym podzielić ten ostatni na dwa: jeden dla salir, drugi dla entrar, a wnioskom nie nadawać statusu rozdziału.

W bibliografii znalazłem usterkę wynikłą zapewne z bezkrytycznego zawierzenia procesorowi tekstów w procesie sortowania. Spis literatury zawiera pozycje przywołane w tekście, co jest zgodne z regulami sztuki. Został on podzielony na kilka części: najpierw znajdujemy listę studiów, artykułów i monografii (w sumie 185 pozycji, dobrze dobranych i dobrze wykorzystanych), dalej słowniki (10 pozycji), źródła internetowe (13 pozycji) i dwie bazy danych. Otoż wydaje się, że ograniczona do adresu pozycja 96 z pierwszej listy (http://www.tdx.cat/bitstream/handle/10803/83983/1g1de1.pdf?jsessionid=84B8CE54CB60CF220E5D57629E5D1A6F.tdx2?sequence=1) powinna zostać połączona z pozycją 86 odsyłającą do pracy doktorskiej obronnej na Universidad Autónoma de Barcelona, (Gyska T., 2011, tesis doctoral disponible en la pagina web). Należałoby też oczekiwać wskazania co najmniej tytułu dysertacji; tego niestety zabrakło.

Każda bibliografia jest odzwierciedleniem treści zawartych w pracy. Otoż czytelnik interesujący się na co dzień tłumaczeniem maszynowym z jednej strony z uznaniem musi przyznać, że Doktorantka panuje nad materiałem z tego zakresu, porusza się po nim z dużą swobodą i proponuje bardzo klarowny przegląd zastanego stanu rzeczy. Z drugiej strony jednak dostrzeże fakty, że swoje uwagi na temat obserwowanych tu trendów Autorka ogranicza w czasie. Trudno czynić z tego poważny zarzut, tak się bowiem składa, że ewolucja w tym zakresie postępuje z szybkością piorunującą, na dodatek zaś ostatnie nowości, które moim zdaniem, mają szansę stanowić znaczący krok na przód, stały się dostępne publicznie zaledwie kilka tygodni temu. Myślę tu o systemie opracowanym wspólnie przez Uniwersytet Harvarda i bardzo doświadczone firmę Systran nazwanym Open-source Neural
Machine Translation System (http://opennmt.net/) System ten, udostępniony o
dziwo z bardzo oszczędnym kodem źródłowym ograniczonym do ok. 4.000 linii (sic!),
oparty jest na idei sieci neuronowych. Pojawia się on po systemach bazujących na
statystyce, z których w pewnym sensie się wywodzi, wykorzystuje także sztuczną
inteligencję i w sumie zdaje się stanowić bardzo obiecującą alternatywę dla
dotychczasowych rozwiązań. Czy więc w kontekście oceny pracy doktorskiej,
pominięcie milczeniem informacji o systemie opartym na sieciach neuronowych
stanowi czynnik umniejszający jej wartość? Nie sądzę, bowiem jest jasne, że w
informatycznej architekturze neuronowej da się z powodzeniem wykorzystać opis
zaproponowanych przez Doktorantkę faktów ścisłe językowych, podobnie jak było to
możliwe w przypadku systemów statystycznych, czy jeszcze starszych systemów
transferowych opartych na regułach.

Opis ten wykonany jest w ramach teoretycznych uwzględniających kilka
elementów: model leksykoograficzny Gastona Grossa, strukturę predykatowo-
argumentową Stanisława Karolaka, teorię Sens-Tekst Igora Mielczuka, słownik w
ujęciu frames i scripts pomysłu James’a Pustejovskiego i Branimira Boguraeva i
wreszcie rozwiązania Xaviera Blanco i jego współpracowników z Barcelony.

Autorka materiał do opisu wybierała z jednej strony z dobrej jakości,
renomowanych słowników, z drugiej zaś wspierała się danymi wydobyтыmi z dwóch
elektronicznych baz danych: Marka Daviesa Corpus del español: 100 million
words, 1200s-1900s (http://www.corpusdelespanol.org/) i Real Academia
Española: Banco de datos CREA (en línea), Corpus de referencia del español
actual http://www.rae.es. Wsparcie korpusowe ułatwiło przygotowanie schematów
syntaktyczno-semantycznych. Punktem odniesienia dla obu leksemów hiszpańskich
jest język polski, dlatego też nie zabrakło źródeł polskich (Perlin O., Perlin J., 2001,
Gran diccionario español-polaco, Warszawa, Wiedza Powszechna i Słownik
kluczowe dla koncepcji G. Grossa pojęcie klas obiektowych. Co za tym idzie wskazane
są czasowniki operatory i typowe dla danej klasy atrybuty. Wyodrębnione w ten
sposób elementy są zebrane w tabelach przeglądowych uwzględniających ogromny
material.

Zaletą pracy jest wyjątkowo drobiazgowo opracowanie schematów składniowo-SEMANTYCZNYCH, które dzięki opozycjom występującym między różnymi występującymi w nich parametrami dają jednoznaczna charakterystykę wybranego
leksemu w języku źródłowym, to zaś z kolei pozwala jednoznacznie przyporządkować odpowiedni leksem w języku docelowym. Sieć tych opozycji jest zaiste imponującą, jej utworzenie zostało dokonane z ogromną precyzją, co wymagało benedyktyńskiej pracy. W efekcie otrzymaliśmy z tekst łączący w sobie oczywisty walor teoretyczny z dużą wartością praktyczną.

Wspomniane powyżej fakty prowadzą do wniosku, że przedłożona mi do oceny praca doktorska mgr Agnieszki Palion-Musiol pt. „Desambiguacion semantica automatica de los verbos seleccionados de movimento salir y entrar en el enfoque orientado a objetos/Automatyczne odwieloznaczenie semantyczne wybranych czasowników ruchu salir i entrar w ujęciu zorientowanym obiektowo” spełnia wymagania stawiane rozprawom doktorskim.

W szczególności stanowi ona oryginalne rozwiązanie ciekawego problemu naukowego mającego nadto poważne implikacje praktyczne, jakimi jest dezambiguizacja leksemów wieloznacznych i stanowi dowód ogólnej wiedzy teoretycznej Doktorantki w zakresie językoznawstwa oraz umiejętności samodzielnego prowadzenia pracy naukowej. Biorąc zatem pod uwagę ustawę z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (por. Dz. U. nr 65, poz. 595, art. 13. 1 z późn. zm.) wnoszę o dopuszczenie jej Autorki do dalszych etapów przewodu doktorskiego.

Prof. dr hab. B. Krzysztof Bogacki
Uniwersytet Warszawski

Warszawa, dnia 15 lutego 2017 roku