

Warszawa, dnia 7.08.2023

prof. ucz. dr hab. Bartłomiej Czaplicki
Instytut Anglistyki
Uniwersytet Warszawski

Ocena rozprawy doktorskiej p. mgr Alony Kononenko-Szoszkiewicz pt. „Exploring (mor)phonotactic patterns: A comparative study of selected Germanic and Slavic languages”

Rozprawa doktorska p. mgr A. Kononenko-Szoszkiewicz składa się z czterech tematycznie powiązanych artykułów naukowych na temat zbitek morfonotaktycznych i fonotaktycznych. Zbiór artykułów jest spójny pod względem treści i celów badań. Analiza zbitek jest wielowątkowa i nowatorska. Z jednej strony analizowane są różnorodne diachroniczne mechanizmy powstawiania zbitek, z drugiej różnice w typach i miejscach występowania zbitek w językach słowiańskich i germańskich. Artykuły prezentują wyniki analiz korpusowych zbitek w językach niemieckim, rosyjskim i ukraińskim. Analizy są profesjonalnie przeprowadzone i dobrze wpisują się w nurt badawczy zapoczątkowany i reprezentowany przez Prof. Wolfganga U. Dresslera i Prof. Katarzynę Dziubalską-Kołaczyk. Nurt ten jest godnie kontynuowany i przez wychowanków wymienionych naukowców, włączając w to p. mgr Kononenko-Szoszkiewicz. Należy szczególnie docenić wyjątkowy kunszt analityczny Doktorantki i mrówczą pracę przy zbieraniu, przygotowywaniu, analizie ogromu danych korpusowych i późniejszej ich interpretacji. Wyniki te nierzadko tylko częściowo wspierają postawione wcześniej hipotezy badawcze, co jest domeną badań empirycznych i przyczynia się do postępu w nauce. Wyniki zaprezentowanych analiz zwiększają naszą wiedzę nie tylko o morfonotaktyce, ale przede wszystkim o akwizycji i przetwarzaniu języka. Wyniki analiz z pewnością przysłużą się następnym badaczom.

Na początku omówię strukturę ocenianej rozprawy doktorskiej. Część pierwsza stanowi wprowadzenie do rozprawy. Definiuje kluczowe pojęcia i przedstawia przyjęte podejścia teoretyczne. W rozdziale tym przedstawiona została analiza bibliometryczna bazy publikacji *Scopus*. Rozdział ten prezentuje też pokrótce główne cele i wnioski wynikające z

przeprowadzonych analiz w czterech artykułach przedstawionych do oceny. Autorka omawia również ograniczenia analiz oraz przyszłe kierunki badań. Na końcu rozdziału przedstawione są główne wnioski wynikające z całej rozprawy doktorskiej.

Artykuł badawczy nr 1 został stworzony przez Doktorantkę oraz pięcioro współautorów. Wkład pracy Doktorantki został oceniony na 15%. Doktorantka była odpowiedzialna za analizę danych z języków ukraińskiego i rosyjskiego. Artykuł ten zawiera wprowadzenie do morfonotaktyki i zwraca szczególną uwagę na różnorodne mechanizmy powstawania morfonotaktycznych zbitek spółgłoskowych. Zawiera ciekawe wnioski na temat różnic w pochodzeniu zbitek w językach słowiańskich i germańskich. Artykuł jest niezwykle przydatnym i wielowątkowym kompendium wiedzy na temat etiologii morfonotaktycznych zbitek spółgłoskowych.

Artykuł badawczy nr 2 ma dwóch autorów. Wkład pracy Doktorantki został oceniony na 60%. Artykuł zawiera porównanie charakterystyki zbitek w językach niemieckim i rosyjskim. Artykuł zawiera ciekawe wnioski. Niemiecki charakteryzuje się występowaniem zbitek na końcu wyrazów, w rosyjskim zbitki występują głównie na początku wyrazów. Ta różnica wynika ze struktury obu języków. Większość morfonotaktycznych zbitek w rosyjskim jest preferowana, co nie zgadza się z postawioną na początku artykułu hipotezą.

Artykuł badawczy nr 3 ma dwóch autorów. Wkład pracy Doktorantki został oceniony na 65%. Artykuł dostarcza typologiczną charakterologię morfonotaktyki i fonotaktyki języka niemieckiego. Artykuł poddaje analizie dane z korpusu, w szczególności skupia się na frekwencji i bada znaczenie type i token frequency dla zbitek. Użycie współczynnika TTR, który bierze pod uwagę zarówno type, jak i token frequency wydaje się zasadne dla badania frekwencji zbitek. Inną zaletą artykułu jest możliwość empirycznej weryfikacji hipotez przy użyciu dużego korpusu. Ten artykuł również podważa hipotezę, że morfonotaktyczne zbitki są mniej preferowane (*more marked*) niż fonotaktyczne zbitki (dla niemieckiego). Artykuł zawiera interesujące typologiczne konkluzje, w szczególności te dotyczące różnic w restrykcjach na peryferyjne spółgłoski w zbitkach: silne dla niemieckiego i angielskiego, słabe dla polskiego i słowackiego.

Artykuł badawczy nr 4. Artykuł ma jedną autorkę. Autorka prezentuje pionierską analizę korpusową fonotaktycznych i morfonotaktycznych zbitek w języku ukraińskim. Artykuł skupia się na zbitkach na początku wyrazów. Autorka wzoruje się na wcześniejszych analizach innych języków. Głównym wnioskiem z analizy jest to, że zbitki morfonotaktyczne są bardziej nacechowane (*more marked*) niż zbitki fonotaktyczne. Artykuł stanowi ważny przyczynek dla dalszych badań nad ukraińskim.

Poniżej prezentuję komentarze, sugestie oraz uwagi polemiczne do dysertacji z podaniem numerów stron.

Introduction

- Str. 12-13: Autorka wskazuje na przewagę zbitek spółgłoskowych na początku wyrazów w języku ukraińskim i rosyjskim, w odróżnieniu od języka angielskiego i niemieckiego, gdzie zaobserwowana jest przewaga zbitek spółgłoskowych w środku i na końcu wyrazów. Autorka wspomina, że różnica ta wynika z różnej specyfiki tworzenia wyrazów złożonych (compounds) w tych językach, jednak nie wychodzi poza to ogólne stwierdzenie. Dyskusja zyskałaby na kilku konkretnych przykładach tworzenia wyrazów złożonych w wymienionych językach.
- Str. 18: Autorka charakteryzuje fonologię Beats-&-Binding (B-&-B) jako “alternative approach” (podejście alternatywne). Brakuje doprecyzowania dla jakiego podejścia B-&-B jest alternatywą. Być może jest to tylko kwestią doboru słownictwa. Jednak na bardziej ogólnym poziomie, rozdział ten zyskałby na prezentacji tradycyjnej teorii sylaby z jej głównymi założeniami i przewidywaniami. W szczególności brak jest odniesienia do zasady regulującej sekwencjonowanie segmentów w sylabie (Sonority Sequencing Principle/Generalization) oraz do tradycyjnej skali sonorności (Jespersen 1904, Selkirk 1986). Ważne byłoby wskazanie na podobieństwa i różnice między tradycyjnym podejściem a teorią B-&-B oraz zwrócenie uwagi na główne argumenty przemawiające na korzyść teorii B-&-B (i ewentualnie jej słabości). Takie szersze spojrzenie umożliwiłoby przedstawienie motywacji dla przyjęcia teorii B-&-B jako bazy teoretycznej dla artykułów składających się na rozprawę doktorską. Rozdział w obecnej formie sugeruje, że teoria B-&-B wyrosła w próżni i nie jest oparta na wcześniejszych badaniach, co nie jest prawdą. Poszerzone informacje na temat tego modelu zawarte są w artykule nr 3 tej dysertacji. Nie zmienia to faktu, że wprowadzenie do dysertacji powinno zawierać takie informacje, gdyż ważne jest umiejscowienie preferowanego modelu w szerszym kontekście i przedstawienie dowodów przemawiających za przyjęciem go do analiz prezentowanych w dysertacji.
- Str.14-15: Opis zasady NAD (*NAD principle*) i sposób obliczania parametrów NAD są momentami zbyt pobieżnie przedstawione. Dla przykładu, nie zostało podane źródło wartości dla parametrów MOA i POA, czyli *Phonotactic Calculator*, Dziubalska-Kołaczyk i in. 2014. Autorka nie podaje też, jakie wartości *NAD product* są preferowane, a jakie nie (tzn. jaki jest punkt odcięcia). Więcej informacji na temat

modelu NAD i jego aplikacji do konkretnych danych zawarte jest dopiero w artykule nr 3 niniejszej dysertacji.

Analiza bibliometryczna

- Str. 20: Dane z Aneksu 2, jak również oparty na nim wykres, Fig. 3, podają frekwencję języków w badaniach korpusowych. Język angielski został odrębnie potraktowany od swoich regionalnych wariantów, np. brytyjskiego, amerykańskiego i australijskiego. Trzeba zauważyć, że termin język angielski jest terminem generycznym i zawiera wszystkie warianty regionalne. Gdyby zsumować frekwencję użycia wszystkich wariantów angielskiego, dominacja tego języka w wynikach byłaby jeszcze wyraźniejsza.
- Nie jest jasne powiązanie tej analizy z tematem rozprawy. Autorka wskazuje na str. 40, że celem tej analizy jest przedstawienie stanu badań i głównych trendów w językoznawstwie korpusowym, jednak nie jest to powiązane z tematem rozprawy. W istocie rzeczy, głównym tematem rozprawy jest badanie wzorców morfonotaktycznych, a analiza korpusowa jest tylko narzędziem tego badania. Celem rozprawy nie jest charakterystyka współczesnych trendów w językoznawstwie korpusowym. Jednakże zarzut ten nie zmienia faktu, że analiza jest dobrze przeprowadzona i zaprezentowana, a wnioski z niej płynące są interesujące. Autorka pokazuje, że analizy korpusowe mogą dostarczyć niezbędnych danych ilościowych do poparcia wniosków na temat kierunków prowadzonych badań w dziedzinie językoznawstwa. Jednak bardziej odpowiednim miejscem dla przedstawienia wyników tych badań byłby odrębny artykuł.

Sekcja *Corpus phonology*:

- W sekcji poświęconej wprowadzeniu metod analizy korpusowej, Autorka nie wyjaśnia kluczowych pojęć: *type* i *token frequency*. Brak też jest wyjaśnienia do czego służą i jak są mierzone (cf. Zydorowicz i in. 2016).

Dyskusja:

- Str. 35: Wyniki badań Autorki podważają wcześniejszą hipotezę, że zbitki morfonotaktyczne są mniej preferowane (tzn. mają bardziej nacechowaną (marked) strukturę) od zbitek fonotaktycznych. Jest to interesujący wniosek, który zasługuje na wyjaśnienie. Dlaczego zbitki morfonotaktyczne nie są stosunkowo bardziej

nacechowane (more marked) niż zbitki fonotaktyczne? Odpowiedź na to pytanie jest szczególnie istotna w świetle artykułu nr 3, gdzie wspomniane jest, że „when a consonant cluster is only morphonotactic, the morpheme boundary is more salient, which facilitates decomposition or segmentation”.

- Czy wynik ten może być spowodowane uniwersalnymi czynnikami wynikającymi z przetwarzania języka (language processing), które mają wpływ na ewolucję języka (w rozumieniu Frischa 2004)? Na przykład można wyobrazić sobie następującą trajektorię. Jest kilka alternatywnych sufiksów tworzących daną kategorię morfosyntaktyczną, ale częściej wybierane są te, które nie tworzą nacechowanych zbitok (bo na przykład zaczynają się od samogłoski). Z czasem te sufiksy stają się produktywne i wypierają te, które tworzą nacechowane zbitki. Innym czynnikiem może być redukcja fonetyczna zbitok (elizja). Być może istnieje inne wyjaśnienie, które Autorka uzna za bardziej przekonujące. Kwestia ta zasługuje na więcej uwagi.
- Str. 36 oraz Typological Conclusions w artykule nr 3

Wniosek, że w niemieckim peryferyjne spółgłoski w zbitkach nie zwiększają nacechowania wewnętrznych spółgłosek w zbitce jest bardzo interesujący i prowadzi Autorów do ciekawych wniosków. W tym kontekście, poniższe zdanie na str. 36 oraz podobne w artykule nr 3 wymagają wyjaśnienia.

“Thus, it seems that in the case of strong restrictions on the selection of the most peripheral consonants, the selection is natural in the sense of not changing the (dis)preferredness of the interior consonant clusters to which they are added.”

W niemieckim to, czy zbitka jest preferowana, czy też nie zależy od peryferyjnych spółgłosek. W polskim peryferyjne spółgłoski mają wpływ na to, czy zbitka jest preferowana, czy nie (*NAD product*). Autorzy wiążą tę różnicę z silnymi restrykcjami (niemiecki) i słabymi restrykcjami (polski) na peryferyjne spółgłoski. Wydaje się, że różnica jest raczej spowodowana tym, że w niemieckim wszystkie peryferyjne spółgłoski są koronalne (przedniojęzykowe). Wcześniejsza literatura wskazuje, że spółgłoski koronalne są mniej nacechowane od spółgłosek o innych miejscach artykulacji (oprócz spółgłosek krtaniowych). Na przykład spółgłoski koronalne podlegają epentezie częściej niż inne, są produktem neutralizacji (choć są też kontrprzykłady), itp. Dla angielskiego, na przykład, została zaproponowana zasada Coronal Coda, która dozwala spółgłoski koronalne nawet gdy pogwałcają preferowaną sekwencję segmentów w sylabie (w ramach SSG), np. w wyrazie *text*. Warto o tym wspomnieć przy omawianiu potencjalnych wyjaśnień dla faktu, że peryferyjne

spółgłoski w niemieckim nie dodają do nacechowania wewnętrznych zbitek. Alternatywnym podejściem, wspomnianym przez Autorów w artykule nr 3, jest odniesienie do ekstrametrykalności, jednak kluczową kwestią pozostaje miejsce artykulacji peryferyjnych spółgłosek.

- Na str. 37 i 41 są sprzeczne wnioski z przeprowadzonych analiz:
„(...) a moderate correlation between the degree of preferredness and frequency can be observed”.
“(...) no significant correlation was observed between cluster preference and its frequency across all languages examined.”
- W tym kontekście istotne jest zwrócenie uwagi na wniosek na str. 37 „the most frequently occurring clusters consist of a mix of both dispreferred and preferred clusters”. Ten wniosek wskazuje na kluczową, a być może nawet nadrzędną nad sonornością rolę frekwencji w przewidywaniu użycia zbitki. Przy dużej frekwencji zbitki, jej struktura pod względem sonorności schodzi na dalszy plan. Warto o tym wspomnieć bardziej dobitnie. Autorka prawidłowo odnosi się w następnym akapicie do badań nad kluczową rolą frekwencji w przetwarzaniu i akwizycji języka, jednak nie wiąże tego jasno z wynikami swoich badań.
- Artykuł nr 3:
- W tabelach brak jest wskazania, czy dane zbitki funkcjonują jako morfonotaktyczne, fonotaktyczne, czy też jako obie te kategorie. Autorzy odnoszą się do tego kryterium wielokrotnie w analizie, więc uwzględnienie tej informacji w omawianych tabelach byłoby wskazane i pomocne.
- Str. 25: W opisie obliczania type-token ratio (TTR) jest nieścisłość. Powinno być:
The number of word types divided by the number of tokens (jest odwrotnie). Dodatkowo brakuje mnożnika (* 100). Natomiast obliczenia są przeprowadzone prawidłowo.
- Wniosek, że morfonotaktyczne zbitki nie są bardziej nacechowane (dispreferred, more marked) od fonotaktycznych zbitek jest bardzo interesujący. Można rozważyć, czy inne sposoby szacowania preferencji zbitek nie dałyby innych rezultatów. Dla przykładu, Autorzy motywują wybór modelu NAD tym, że oferuje „the broadest existing possibility for defining degrees of segmental cohesion” na str. 18. Należałoby

sprawdzić, czy przewidywania tego modelu nie są zbyt restrykcyjne w porównaniu z innymi modelami. Porównanie przewidywań modeli NAD i SSG dla polskich zbitek zostało wykonane w Zydorowicz i Orzechowska (2017). Model mniej restrykcyjny okazał się bardziej dopasowany do danych. Autorzy wspominają też o pracach Orzechowska i Wiese (2011, 2015), gdzie preferencje zbitek były badane podejściem empirycznym przy użyciu cech fonologicznych. Wskazuje to na możliwość istnienia różnic między językami nie wynikających jedynie z uniwersaliów opartych na sonorności (SSG, NAD), ale również z czynników związanych ze specyfiką rozwoju danego języka. Warto o tym wspomnieć.

- Artykuł nr 4
- Opisy przy Tabelach 3 i 4 błędnie sugerują, że zawierają one dane na temat *word-type frequency*. Takie dane są zawarte w aneksach, nie w tabelach w artykule.
- Wyniki analiz regresji liniowej nie są wystarczająco dokładnie opisane. Na przykład, dla zbitek dwuspółgłoskowych Autorka wskazuje, że jest znacząca statystycznie zależność między zmiennymi *NAD product* i *FreqLog* dla preferowanych zbitek. Jednak po bliższym przyjrzeniu się pierwszemu wykresowi na Figure 2 wydaje się, że ta zależność (korelacja) jest odwrotna niż przewidywana. Im wyższa wartość *NAD product*, tym niższa *FreqLog*. Negatywna korelacja jest obserwowana również dla wszystkich razem (preferowanych i nie) zbitek. Taki wynik sugeruje istnienie tendencji, że im lepsze zbitki są pod względem NAD, tym rzadziej występują. Nie ma o tym wzmianki w opisie wyników. Dla zbitek trójspółgłoskowych wyniki regresji liniowej są bardziej zgodne z przewidywaniami (pozytywna korelacja, choć dość niska). Ciekawe byłoby też poznać sumaryczne wyniki regresji liniowej dla wszystkich badanych zbitek, zarówno dwu- jak i trójspółgłoskowych. Zbitek dwuspółgłoskowych jest znacznie więcej niż zbitek trójspółgłoskowych, co oznacza, że wykryta negatywna korelacja dla zbitek dwuspółgłoskowych może mieć znaczący wpływ na całość danych i na implikacje teoretyczne. Implikacje wynikające z tych bardziej poszerzonych wyników są kluczowe i powinny być uwzględnione w podsumowaniu artykułu.
- Użyte formuły i całosciowe wyniki analiz statystycznych powinny być przedstawione w aneksie dla większej przejrzystości.

Praca jest napisana bardzo dobrą angielszczyzną akademicką (nie oceniam tu artykułów współautorskich, bo kwestie poprawności językowej zależały od wszystkich autorów), jednak zdarzają się drobne usterki. Na przykład na str. 24 powinno być „as **has** been demonstrated”, na str. 26 powinno być “Romance languages”, zamiast “Roman languages”, na str. 27 powinno być „All consonant clusters were divided according to **their** status”, na stronie 28 jest “the main finding of our research goes **against with** the claim” (?), na str. 32 powinno być “Ukrainian **tends** to avoid (...) and usually **creates** ...”. Na str. 32 jest sprzeczność: “adding another phoneme to the **word coda** increases cluster complexity”. W omawianym artykule były badane zbitki **na początku** wyrazów, nie na końcu.

Na ocenianą serię artykułów składają się jeden artykuł autorski oraz trzy współautorskie (wkład pracy mgr Kononenko-Szoszkiewicz 15%, 60%, 65%). Muszę zauważyć, że ocena rozprawy na podstawie publikacji współautorskich jest utrudniona, bo nie zawsze jest jasne, które fragmenty analiz empirycznych i teoretycznych można przypisać Doktorantce. Jednak na korzyść p. mgr Kononenko-Szoszkiewicz przemawia kilka argumentów: (i) współautorzy są uznanymi naukowcami, którzy podjęli się współpracy z Doktorantką, (ii), jeden ze współautorów, Profesor Dressler, jest też jej promotorem, (iii) widać stopniowy postęp w deklarowanym wkładzie pracy Doktorantki w artykuły, co wskazuje na wzrost samodzielności w przeprowadzaniu analiz i interpretacji wyników, (iv) wspólnie normą jest, że prace zawierające badania empiryczne powstają we współautorstwie, (v) publikacje ukazały się (bądź ukażą się) w prestiżowych wydawnictwach i czasopismach.

Powyższe uwagi nie wpływają na moją zasadniczo pozytywną ocenę recenzowanej rozprawy doktorskiej. Przedstawione wyniki badań znacząco poszerzają naszą wiedzę na temat funkcjonowania i przetwarzania języka, a w szczególności powstawania i charakterystyki zbitek spółgłoskowych w językach germańskich i słowiańskich. Zebrane dane empiryczne i ich analizy stanowią cenny wkład w rozwój dyscypliny i stanowią oryginalne rozwiązanie problemu naukowego. Doktorantka ma szeroką wiedzę teoretyczną w dyscyplinie językoznawstwo i potrafi samodzielnie prowadzić pracę naukową. W konkluzji z **pełnym przekonaniem stwierdzam, iż przedstawiona do oceny przez p. mgr Alony Kononenko-Szoszkiewicz rozprawa doktorska pt. „Exploring (mor)phonotactic patterns: A comparative study of selected Germanic and Slavic languages” spełnia wymogi stawiane**

pracom doktorskim i wnioskuje o dopuszczenie Doktorantki do dalszych etapów postępowania doktorskiego.

Warszawa, 7 sierpnia 2023 roku

A handwritten signature in blue ink, reading "Bartłomiej Czaplicki". The signature is written in a cursive style with a horizontal line through the middle of the name.

Prof. ucz. dr hab. Bartłomiej Czaplicki