

## Autoreferat

1. **Imię i Nazwisko:** Beata Wójtowicz
2. **Posiadane dyplomy, stopnie naukowe – z podaniem nazwy, miejsca i roku ich uzyskania oraz tytułu rozprawy doktorskiej.**

2000 rok – Uniwersytet Warszawski, Wydział Neofilologii, Instytut Orientalistyczny, tytuł magistra kulturoznawstwa, specjalność afrykanistyka uzyskany na podstawie pracy magisterskiej pt. „Czasowniki posiłkowe w suahili na tle procesu gramatyzacji” napisanej pod kierunkiem dr Iwony Kraski-Szlenk;

2005 rok – Uniwersytet Warszawski, Instytut Orientalistyczny, stopień doktora nauk humanistycznych w zakresie językoznawstwa na podstawie rozprawy doktorskiej pt. „Ogólnolingwistyczne i leksykograficzne podstawy organizacji słownika suahilijsko-polskiego” napisanej pod kierunkiem prof. dr. hab. Janusza S. Bienia.

3. **Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych.**

2005 – obecnie: adiunkt w Katedrze Języków i Kultur Afryki Wydziału Orientalistycznego Uniwersytetu Warszawskiego,

2012 – 2013: lingwista w Instytucie Podstaw Informatyki Polskiej Akademii Nauk; z Instytutem współpracuję nieustająco na podstawie umów o dzieło/zlecenie od roku 2001,

2000 – 2005: studia doktoranckie na Wydziale Neofilologii Uniwersytetu Warszawskiego.

4. **Wskazanie osiągnięcia\* wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2016 r. poz. 882 ze zm. w Dz. U. z 2016 r. poz. 1311.):**

a) tytuł osiągnięcia naukowego:

**Leksykografia suahili: słowniki i ich użytkownicy.**

b) (autor/autorzy, tytuł/tytuły publikacji, rok wydania, nazwa wydawnictwa, recenzenci wydawniczy),

***Monografia:***

Wójtowicz Beata, *Swahili Lexicography. Modern Dictionaries and their Users. A Swahili-Polish Dictionary Case Study*, 2018, Lincom Europa.

Recenzent wydawniczy: prof. D. J. Prinsloo, *Department of African Languages, University of Pretoria, RPA.*

***Artykuły:***

1. Wójtowicz Beata, Evaluating a 12-Million-Word Corpus as a Source of Dictionary Data, 2017, *International Journal of Lexicography*, doi: 10.1093/ijl/ecx011, publikacja dostępna online od 03.08.2017. (czasopismo z bazy Web of Science)
2. Wójtowicz Beata, Revisiting Lemma Lists in Swahili Dictionaries, 2017, *Lexikos 27*, s. 561-577. (czasopismo z bazy Web of Science)
3. Wójtowicz Beata, Learner Features in a New Corpus-based Swahili Dictionary, 2016, *Lexikos 26*, s. 402-415. (czasopismo z bazy Web of Science)

c) omówienie celu naukowego ww. prac i osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania.

Wskazana, jako osiągnięcie naukowo-badawcze, monografia wieńczy na obecnym etapie moje wieloletnie badania nad leksykografią języka suahili. W osiągnięciu wykazuję również poprzedzające tę monografię publikacje, które omawiają badania, do których się odnosi i dalej je

rozwija (pkt. 1 i 2) oraz słownik, którego badania te głównie dotyczą (pkt. 3).

Moja działalność badawcza sytuuje się w dziedzinie metaleksykografii i dotyczy metod opracowywania słowników, badań nad ich użyciem oraz historii leksykografii.

Głównym celem badań przedstawionych w publikacjach jest pogłębienie wiedzy na temat użycia słowników języka suahili (i polskiego) oraz ewaluacja zarówno założeń metodologicznych prezentacji materiału słownikowego, jak i wykorzystania źródeł dostępnych w postaci elektronicznych korpusów tekstów w pracach leksykograficznych. Celem pośrednim jest również zbadanie stopnia wykorzystania nowoczesnych metod leksykografii światowej w produkcji słowników dla języków mniej rozwiniętych pod względem dostępności nowoczesnych narzędzi informatycznych i źródeł elektronicznych, w tym suahili, ogólniej bantu i innych języków afrykańskich.

Słowniki stanowią podstawowe narzędzie w poznawaniu języków obcych. Są produktami tworzonymi z myślą o nabywcy i z założenia mają zaspokajać jego potrzeby. Takie pojmowanie leksykografii jest powszechne w kulturze zachodniej, gdzie wydawanie słowników było od dawna przedsięwzięciem bardziej komercyjnym niż, jak np. w tradycji polskiej, powinnością narodową. Obecna komercjalizacja leksykografii ma również wpływ na jej nowe postrzeganie i definiowanie. Sama leksykografia ma jednak nadal dwa wymiary: poza praktycznym, również naukowy, teoretyczny, odnoszący się do metod i technik opracowywania słowników i tego aspektu dotyczą moje badania.

Słowniki języków europejskich, w szczególności języka angielskiego, są szeroko badane pod różnym kątem, a wyniki tych badań są wykorzystywane w ich dalszym rozwijaniu. Forma elektroniczna i dostępność elektronicznych korpusów tekstów wprowadziły natomiast szereg nowych możliwości, które przyczyniły się do opracowania nowych dzieł, które technologiczne nowinki wykorzystują w celu lepszej organizacji i prezentacji materiału i dostosowaniu treści do wymagań i oczekiwań konkretnych odbiorców. Słowniki języków mniej popularnych, rzadko nauczanych, o mniejszych zasobach z takich możliwości korzystają rzadziej, rzadko też, o ile w ogóle, poddawane są badaniom.

Język suahili jest jednym z najlepiej znanych i opisanych języków afrykańskich a mimo to należy zaliczyć go do grupy języków o małych zasobach (*less-resourced languages*), jeśli chodzi o narzędzia informatyczne i korpusy tekstów elektronicznych. Dla suahili, jako jednego z zaledwie kilku języków afrykańskich, został opracowany anotowany korpus elektroniczny, daje on jednak dostęp do zaledwie 20 milionów słów. Jak dotąd jest on sporadycznie wykorzystywany w badaniach nad językiem, a prawie zupełnie w leksykografii. Suahili, choć ma wielomilionową grupę użytkowników, jest w grupie języków rzadko nauczanych (*less-commonly-taught languages*) w

Europie, Ameryce, jak również w Azji. Brakuje zarówno podręczników do nauki języka, jak i słowników, które opracowywane są w zasadzie jedynie dla języka angielskiego. Brakuje też badań teoretycznych nad nauczaniem i leksykografią języka. Niewykorzystane pozostają również możliwości, jakie oferuje forma elektroniczna słowników. Większość nowych publikacji nadal opiera się na dziełach wcześniejszych i publikowanych jest w formie tradycyjnej, a ich statyczne kopie elektroniczne stanowią jedynie załącznik do papierowych publikacji. Swoimi badaniami staram się zmienić ten stan rzeczy i pokazać, że nawet niewielkie zasoby mogą być wykorzystane w tworzeniu nowoczesnej leksykografii dla języków mało popularnych, a forma elektroniczna daje nowe, wciąż mało doceniane i wykorzystywane możliwości, które mogą zmienić oblicze słowników.

Monografia *Swahili Lexicography. Modern Dictionaries and their Users. A Swahili-Polish Dictionary Case Study* składa się z czterech części, w których omówione zostały metody współczesnej leksykografii, historia leksykografii suahili i badania nad użyciem słowników.

Część pierwsza: *Modern methods of lexicography and their application to lesser-resourced languages* prezentuje osiągnięcia współczesnej leksykografii. Po pierwsze omówione zostały typy słowników i ich funkcje, ze szczególnym uwzględnieniem słowników dwujęzycznych przeznaczonych dla uczących się języków obcych i słowników języków mniej popularnych. Słowniki takich języków trudno jest opracowywać w zgodzie z zasadami współczesnej leksykografii, gdzie nadrzędnym czynnikiem decydującym o strukturze i zawartości leksykonu jest sprecyzowana grupa odbiorców lub konkretna jego funkcja. Ograniczenia zarówno osobowe, jak i finansowe skłaniają raczej do tworzenia dzieł ogólnych, przeznaczonych dla szerokiego odbiorcy i mających na celu w większym stopniu dokumentację języka niż zaspokojenie konkretnej potrzeby potencjalnego użytkownika. W publikacji poruszone zostały również zagadnienia związane z wykorzystaniem korpusów tekstów w pracy leksykografa. Ich przydatność nie jest dziś kwestionowana, jednakże współczesne korpusy dla języków europejskich osiągnęły imponujące rozmiary, zupełnie nieosiągalne dla języków np. afrykańskich. Omawiana zatem jest kwestia, czy wobec braku dużych korpusów, małe ich wersje mają prawo bytu we współczesnych realiach. Pod uwagę wzięte zostały również takie zagadnienia, jak zrównoważenie i reprezentatywność korpusów, a także zaprezentowane zostały różne inicjatywy tworzenia korpusów języków rzadkich, afrykańskich.

Druga część monografii dotyczy leksykografii języka suahili. Prezentuję tu moje wcześniejsze badania, które tworzą tło dla badań nad użyciem słowników i ukazują wiedzę na temat dotychczasowych dokonań leksykografii suahili. Część ta jest niejako kompendium wiedzy o trendach w kompilacji słowników języka. Po krótkim przedstawieniu samego języka i kwestii

gramatycznych, istotnych dla tematu, przechodzę do historii tworzenia słowników dostępnych dla języka suahili opracowywanych od drugiej połowy XIX wieku. W opracowaniu uwzględniłam zarówno kwestie historyczne, które miały wpływ na kształt suahilijskiej leksykografii, jak i słowniki współczesne, jak dotąd pomijane w różnych, zwykle wrywkowych opracowaniach na ten temat. Literatura poświęcona słownikom języka suahili skupia się przede wszystkim na jednojęzycznym słowniku wydawnictwa TUKI (opublikowanym w 1981) i jego późniejszych edycjach, żywo dyskutowanych wśród rodzimych leksykografów. Poruszane kwestie dotyczą głównie zasad konstruowania jednojęzycznych definicji i uwzględniania wariantów, a tym samym zasad tworzenia i roli słownika w opracowywaniu i rozpowszechnianiu języka standard, kwestii ważnych również dla leksykografii dwujęzycznej. W części tej, na podstawie analizy głównie dwujęzycznych słowników języka suahili, przedstawione są wybrane zagadnienia odnoszące się zarówno do makro, jak i mikrostruktury słowników.

Jednym z zagadnień jest sposób hasłowania wyrazów w słowniku. Suahili jest językiem bantu i z typologicznego punktu widzenia wyróżnia się w nim rodzaj gramatyczny, zwyczajowo nazywany klasami. Języki te charakteryzuje aglutynacja, związek zgody, rozwinięta koniugacja prefiksalna, powszechne użycie form lokatywnych i derywacja jako najbardziej produktywny sposób wzbogacania słownictwa. Aglutynacja najwyraźniej widoczna jest w morfologii czasownika. Jego jądro stanowi niepodzielny rdzeń, do którego dołączane są różne morfemy gramatyczne i derywacyjne w pozycji prefiksalnej i sufiksalnej. W morfologii rzeczowników aglutynację zaobserwujemy w rzeczownikach odczasownikowych, które zbudowane są na bazie tematów czasownikowych, tj. rdzeni wraz z dołączonymi sufiksami. Występowanie klas nominalnych i bezpośrednio z nimi związanego związku zgody jest najbardziej charakterystyczną cechą języków bantu. Klasy nominalne wyrażone są morfologicznie, poprzez odpowiednie prefiksy, oraz syntaktycznie – poprzez kongruencję w obrębie zdania. Podstawową operacją słowotwórczą jest derywacja poprzez sufiksację. Rodziny słowotwórcze w językach bantu nierzadko przekraczają sto leksemów. Rzeczowniki odczasownikowe mogą być tworzone za pomocą pięciu różnych procesów derywacyjnych, które opierają się na kombinacji sufiksów derywacyjnych i prefiksów nominalnych. Taka odmienność typologiczna języka wprowadza szereg problemów niespotykanych w nauczaniu i leksykografii języków europejskich. Z perspektywy leksykograficznej dotyczą one przede wszystkim wyboru postaci wyrazu hasłowego oraz sposobu traktowania i prezentacji derywatów oraz zawartości komponentu gramatycznego artykułów hasłowych. Inaczej niż dla np. języka polskiego, suahili nie ma wypracowanych zasad hasłowania, które byłyby jednorodnie stosowane w słownikach. Zasady te różnią się i użytkownik musi poznać je w przypadku każdego słownika, z którego korzysta. Dotyczy to samego hasłowania, gdzie różnice są szczególnie

widoczne w postaci zaimków ale również zasad traktowania derywatów. Jedno z podejść zakłada nadawanie derywatom statusu haseł głównych, drugie opis derywatów pod ich rdzeniami. Trzecia metoda to podejście mieszane, w którym część derywatów ma status haseł, a część opisana jest w ramach hasła rdzenia. Prezentowany w książce postulat zakłada nadanie wszystkim derywatom statusu haseł głównych przy jednoczesnym zachowaniu wielopoziomowych relacji między nimi. Jest to podyktowane głównie względami dydaktycznymi, gdyż uczący się języka często nie są w stanie rozpoznać derywatu i właściwie powiązać go z rdzeniem. Wersja drukowana, ze względu na formę i ograniczenia objętości, najczęściej nie daje możliwości zachowania informacji o derywacji. Forma elektroniczna natomiast pozwala na jawne zachowanie relacji i wizualizację ich np. w formie drzewa derywacyjnego, jak to zostało przedstawione w słowniku suahili-polskim mojego autorstwa. Implementacja struktury danych słownika w języku XML dała możliwość uniknięcia konieczności wyboru pomiędzy powyższymi skrajnymi stanowiskami i pozostawienia go użytkownikowi, który ma możliwość wyboru pomiędzy trybem, w którym cała rodzina derywatów czasownikowych jest zgrupowana w jednym haśle, bądź rozdzielona na samodzielne hasła powiązane jedynie odnośnikami hipertekstowymi. Osiągnięcie to niejako zamyka debatę, pokazując wyższość formatu elektronicznego nad drukowanym. Ze względu na dydaktyczny charakter słownika, wielopoziomowość derywacji została również zachowana i unaoczniona użytkownikowi w postaci graficznego drzewa. Bezpośrednio w artykułach hasłowych podana została informacja o derywatach czasownikowych – formach pochodnych czasownika, drzewo natomiast prezentuje całą rodzinę derywacyjną, uwzględniając również inne części mowy. Podejście takie nie jest jeszcze stosowane w żadnym innym słowniku języka suahili.

Zgodnie ze sztuką nowoczesnej leksykografii wszystkie decyzje leksykograficzne powinny być podyktowane przez funkcje, jakie słownik ma pełnić dla różnych grup jego odbiorców. Słowniki mają tyle celów, ile grup użytkowników. Od grupy odbiorców uzależnia się zarówno makrostrukturę, jak i mikrostrukturę słownika. Żeby móc temu sprostać, powinniśmy znać potencjalnego użytkownika. Przed przystąpieniem do opracowywania podstawowych założeń słownika trzeba odpowiedzieć na kluczowe pytania: kto będzie użytkownikiem słownika, w jakim celu będzie go używał, jakimi umiejętnościami, wiedzą i znajomością języka dysponuje oraz, czego konkretnie od słownika oczekuje. Dobrze, gdy odpowiedzi na powyższe pytania są poparte badaniami rynku.

Badania rynku bywają przeprowadzane przez dużych wydawców, ich wyniki nie są jednak z reguły publikowane. Wiele słowników, szczególnie dla języków rzadko nauczanych, jest opracowywanych poza wydawnictwami, a ich autorzy swoje decyzje opierają często na dostępnych badaniach naukowych dla innych języków.

Odniesienia do użytkowników i ich potrzeb pojawiały się w dziełach leksykograficznych od zawsze. Dopiero jednak w wieku XX opracowano założenia teoretyczne, które stanowią podstawę naukową współczesnych słowników i pojawiła się głębsza refleksja nad funkcją, którą słowniki pełnią. Badania nad użyciem słowników stanowią najmłodszy i obecnie najbardziej rozwijający się dział metaleksykografii. Badanie zachowań i oczekiwań użytkowników słowników dla języków popularnych są prowadzone na szeroką skalę w wielu ośrodkach uniwersyteckich świata. Pierwsze systematyczne badania zostały rozpoczęte nie tak dawno, bo w latach 60. XX wieku i dotyczyły głównie słowników języka angielskiego, oraz innych popularnych języków europejskich, jak francuski czy niemiecki. Bada się zarówno użytkowników słowników jedno, jak i dwujęzycznych, przeznaczonych dla rodzimych użytkowników języka, jak i dla uczących się. Prym w badaniach wiodą badacze języka angielskiego, dla którego powstały najbardziej zaawansowane słowniki, które mają też największą grupę odbiorców na całym świecie. Zaangażowani w nie są nie tylko sami naukowcy, ale również stojące za nimi wydawnictwa, jak np. Macmillan czy Oxford.

Współczesne badania coraz częściej skupiają się na konkretnej funkcjonalności słownika, i wykorzystują możliwości oferowane przez komputery, jak np. okulografia.

Szeroka literatura na temat użycia słowników angielskich wniosła wiele do leksykografii teoretycznej. Wiemy np. że główną informacją poszukiwaną przez większość użytkowników jest znaczenie słów. Dla Włochów uczących się języka angielskiego jako obcego, w pierwszym etapie nauki najważniejsza jest jednak wymowa. Tego typu informacje są wykorzystywane przez leksykografów pracujących z różnymi językami. Nadal brakuje jednak badań szczegółowych nad specyfiką użycia słowników języków rzadko nauczanych, które zwykle typologicznie i kulturowo są bardzo odległe od języka i kręgu kulturowego osób uczących się ich.

Badania nad użyciem słowników są głównym tematem prezentowanej monografii. W części trzeciej: *Research into dictionary use* omawiane są ogólne założenia badań nad użytkownikami i użyciem słowników, ich metody i wybrane wyniki. Kwerenda bibliograficzna w Tanzanii umożliwiła mi dotarcie do badań przeprowadzonych dla języków afrykańskich, w tym suahili, opisanych w niepublikowanych pracach magisterskich i doktorskich Uniwersytetu w Dar es Salaam. Szczególną uwagę zwracam na metodę analizy rejestrów zapytań (*log files analysis*), która została także zastosowana w moich własnych badaniach nad elektronicznym słownikiem suahili-polskim. Badania te szczegółowo prezentowane są w części czwartej: *Case study: Swahili-Polish dictionary user research*.

Badania nad słownikami języków afrykańskich praktycznie nie istnieją. Dlatego nawet badania prowadzone na małą skalę, w oparciu o niewielką liczbę użytkowników/respondentów czy nad jednym wybranym aspektem są cenne. Im więcej takich badań zostanie przeprowadzonych i

upublicznionych, tym pełniejszy obraz leksykografii będziemy w stanie wyłonić. Nie można się spodziewać, że dla języków rzadkich badania osiągną skalę porównywalną do tej dla języka angielskiego, tak jak nie stało się tak w przypadku kompilacji korpusów. W monografii prezentuję badania nad bardzo rzadką parą językową, jaką jest polski i suahili, z nadzieją na pogłębienie wiedzy zarówno o leksykografii polskiej, suahilijskiej, jak i języków bantu w ogóle.

Badania nad słownikami języka suahili prowadzę od wielu lat. Pierwszy raz, prawie 10 lat temu, przeprowadziłam pierwszy kwestionariusz wśród uczących się suahili na Uniwersytecie Warszawskim. Ten sam kwestionariusz został następnie przedłożony studentom kilka lat później. Nawet tak proste badanie wykazało zmiany w postrzeganiu i użytkowaniu słowników związane przede wszystkim z rozwojem technologii i w dostępie do Internetu. Jeszcze 10 lat temu wszystkie badane osoby posiadały i korzystały ze słowników drukowanych. Badanie przeprowadzone niedawno wykazało, że ok. 30% użytkowników nie posiada już w ogóle słowników tradycyjnych, natomiast wszyscy korzystają ze słowników elektronicznych. Wśród najlepiej ocenianych znalazł się nowy słownik suahili-polski, który w innym badaniu został jednak oceniony jako niewystarczający dla wszystkich poziomów, ze względu na mały zasób słownictwa oraz nieprzystosowanie do obsługi na telefonach komórkowych. Użytkownicy najczęściej poszukują informacji o znaczeniu wyrazu, potem o ekwiwalencie, części mowy i klasie rzeczownika. Wysoko oceniają też dostępność przykładów użycia. Tego typu informacje są bardzo przydatne w planowaniu struktury opisu haseł i wskazują elementy, które powinny zostać opracowane z większą starannością.

Głównym obiektem prezentowanych badań jest nowy słownik suahili-polski opublikowany w Internecie pod koniec 2012 roku ([www.kamusi.pl](http://www.kamusi.pl)). Przez okres czterech lat rejestrowane były zapytania wysyłane przez użytkowników do słownika. Następnie na podstawie analizy ponad 50 tysięcy haseł-zapytań ewaluacji została poddana struktura słownika, przyjęte zasady hasłowania i traktowania derywatów. Pod uwagę zostały wzięte zarówno dane suahili, jak i polskie.

Analiza danych suahili wykazała, że struktura kategorii wyrazów poszukiwanych przez użytkowników odpowiada strukturze słownika. Czasowniki są kategorią poszukiwaną najczęściej, jednak to rzeczowniki stanowią największą grupę wśród szukanych wyrazów. Wyrazy funkcyjne, często używane, wbrew oczekiwaniom są również często sprawdzane. Ważną grupę stanowią również pełne formy ortograficzne, jak np. rzeczowniki w liczbie mnogiej, przymiotniki z prefiksem klasowym czy zaimki, które nie były uwzględniane w słownikach do tej pory. Z danych wynika też, które klasy są szczególnie w polu zainteresowań użytkowników. Analizowane dane jasno pokazują, że nadanie derywatom statusu haseł głównych to ukłon w stronę użytkowników. Na ich podstawie jesteśmy w stanie określić, które czasowniki, oraz które ich formy pochodne są



poszukiwane najczęściej i powinny na pewno być uwzględniane w słownikach, a być może także w większym wymiarze w sylabusach nauczania języka.

Zapytania zostały także przeanalizowane pod względem dziedzin tematycznych (*semantic domains*). Spośród dziewięciu dziedzin branych pod uwagę, dwie okazały się być najbardziej popularne, tj. czynności (*physical actions*) i stany (*states*).

Analiza zapytań, które nie zostały znalezione w słowniku pozwoliła na kategoryzację błędów, które popełniają użytkownicy. Jednym z najczęstszych jest wpisywanie w okienko wyszukiwania pełnych form wyrazowych zamiast form hasłowych. Użytkownicy zdają się obecnie traktować słowniki porównywalnie do wyszukiwarek internetowych typu *Google* i zamiast form hasłowych wpisują odmienione wyrazy, frazy a nawet całe zdania. Wykazały to zarówno dane suahili, jak i polskie.

Kolejnym poruszonym zagadnieniem jest korelacja pomiędzy wyszukiwaniami a danymi korpusowymi. Zgodnie z metodą współczesnej leksykografii lista haseł słownika powinna być opracowana na podstawie listy frekwencyjnej otrzymanej z korpusu. Badanie ma na celu analizę, czy użytkownicy rzeczywiście szukają częstych słów, a tym samym, czy lista frekwencyjna to dobre źródło danych dla słownika. Jest to zagadnienie interesujące tym bardziej, że korpus na podstawie którego mogą być tworzone słowniki języka suahili nie jest reprezentatywny i zawiera stosunkowo małą ilość słów. Analizowane dane potwierdzają, że użytkownicy najczęściej szukają częstych słów i w zasadzie prawie wszystkie ich zapytania dotyczą słów, które znajdują się również na liście frekwencyjnej 20 milionowego korpusu. Jest to informacja bardzo cenna w przypadku języków, dla których nie ma wielu zasobów, a słowniki są małe lub średniej wielkości. Opracowanie takiego słownika, nawet na podstawie danych z niewielkiego korpusu, powinno spełnić oczekiwania większości użytkowników. Porównywane także są różne korpusy języka suahili, zarówno korpus opracowany ręcznie, jak i kompilowany automatycznie z danych dostępnych w Internecie (*web-crawled corpus*). W zestawieniu znalazły się również dwie wersje korpusu języka polskiego, które znacząco różnią się wielkością (1 milion i 300 milionów słów).

Swoją pracą staram się wypełnić lukę w badaniach nad słownikami języka suahili, które mogą też być przydatne dla innych języków bantu. Ze względu na strukturę języków bantu, metody wypracowane dla języków europejskich, a głównie angielskiego, nie zawsze mogą być bezkrytycznie zastosowane w słownikach tych języków. Z drugiej strony metody te, szczególnie nowoczesne wykorzystanie narzędzi informatycznych, korpusów tekstów czy formatu elektronicznego, mogą wpłynąć na jakość słowników i pomóc w rozwiązaniu szeregu problemów, z którymi boryka się leksykografia języków bantu od początku swego istnienia, jak np. wspomniany problem hasłowania.

## **5. Omówienie pozostałych osiągnięć naukowo - badawczych.**

Mój rozwój naukowy wiąże się przede wszystkim z leksykografią języka suahili i wykorzystaniem nowoczesnych technologii językowych w pracy językoznawcy i leksykografa. Zainteresowanie przydatnością i rolą słowników w nauce języków obcych oraz adekwatnością ich zawartości do oczekiwań użytkowników, skłoniło mnie do wieloletnich badań. Większość badań prowadzonych jest z perspektywą ich praktycznego zastosowania. Najpierw badałam słowniki już dostępne, ich historię i strukturę, a następnie zajęłam się opracowaniem nowego słownika w oparciu o nowoczesne założenia leksykografii światowej. Dopiero opracowanie własnego słownika pozwoliło mi na prowadzenie badań nad jego użyciem w oparciu o autentyczne dane bez potrzeby sięgania po subiektywne oceny samych użytkowników. Każdy z etapów wymagał kilku lat pracy, trzy lata to jedynie okres trwania projektu, w ramach którego powstawał słownik, kolejne cztery lata, to okres zbierania danych do dalszych badań. W tym czasie moja praca naukowa wiązała się też z innymi projektami. Poniżej pokrótce przedstawiam dziedziny, w których badania jestem zaangażowana. Dziedziny te z oczywistych względów często przenikają się i mają punkty wspólne.

### **5.1 Leksykografia suahili**

Za jedno z najważniejszych osiągnięć uznaję publikację nowego słownika suahili-polskiego w wersji elektronicznej oraz następnie opracowanej na jego podstawie tradycyjnej publikacji w wersji papierowej („Elektroniczny słownik suahili-polski: [www.kamusi.pl](http://www.kamusi.pl)” i „Słownik suahili-polski”). W szeregu artykułów poprzedzających publikację nowego słownika przedstawiałam rozważania na temat jego planowanej struktury. W artykułach skupiałam się na jego strukturze oraz wersji elektronicznej i możliwościach, które taki format oferuje wraz z formatem opisu danych XML („The Logical Structure of Swahili-Polish Dictionary Entries and its XML Implementation”, „A Concept of a New Swahili-Polish Dictionary”, „Multi-level reference hierarchies in a dictionary of Swahili”, „New XML-encoded Swahili-Polish Dictionary: Micro- and Macrostructure”). Prace nad słownikiem były finansowane ze środków grantu z KBN, którego byłam kierownikiem.

W ramach zainteresowań leksykograficznych zajmowałam się również historią leksykografii języka suahili, która nie jest dobrze opisana, a w którą duży wkład miał polski uczonec prof. Rajmund Ohly – wieloletni pracownik Uniwersytetu w Dar es Salaam i redaktor pierwszego jednojęzycznego słownika suahili wydanego w Tanzanii. Wyniki swoich badań opisuję w kilku artykułach (m.in. „Polskie akcenty w historii leksykografii suahili”, „A glimpse at the history of Swahili lexicography”).

## 5.2 Formaty danych

Równolegle prowadziłam prace nad innymi słownikami (suahili-polskim i suahili-angielskim) i eksperymentowałam z formatami danych w ramach projektu *Free Bilingual Dictionaries FreeDict*. Projekt ten ma na celu darmowe udostępnianie słowników i infrastruktury w Internecie na otwartych licencjach. W swoich pracach (prowadzonych z dr. P. Bańskim) testowałam różne formaty danych i opracowałam słowniki w formatach DICT oraz XML zgodnym z TEI. „Swahili-English xFried/Freedict Dictionary 0.3.” to pierwszy słownik w projekcie *FreeDict* opracowany w zgodzie ze standardem TEI P5. Na jego podstawie została opracowana struktura danych dla całego projektu i dokonano konwersji słowników dla innych języków. Zagadnienia związane z formalizmami danych zostały poruszone w artykule „A Repository of Free Lexical Resources for African Languages: The Project and the Method” oraz omówione w kilku wystąpieniach konferencyjnych, m.in. podczas spotkania konsorcjum TEI: „Freedict: an Open Source repository of TEI-encoded bilingual dictionaries”.

## 5.3 Rola licencji autorskich w tworzeniu zasobów

Interesuje mnie również zagadnienie wolnego dostępu do danych i narzędzi informatycznych. Takie podejście do zasobów jest bardzo korzystne dla języków, które mają ich niewielką ilość. Wymiana danych i współpraca różnych twórców przyczynia się do tworzenia nowych zasobów, co jest szeroko udokumentowane. Przykładem tego może być *Swahili Verb Analyser*<sup>1</sup>, darmowy analizator czasowników suahili, który został opracowany na podstawie słownika mojego autorstwa udostępnionego w ramach projektu *FreeDict* („Swahili-English English-Swahili xFried Dictionary”). Tematyka związana z potrzebą dostępności darmowych narzędzi i źródeł była poruszana w artykułach („A Repository of Free Lexical Resources for African Languages: The Project and the Method”, „Learner Features in a New Corpus-based Swahili Dictionary”) oraz w kilku wystąpieniach konferencyjnych („Free bilingual dictionaries with a focus on Swahili: the state of affairs and the prospects for the future”, „Lexicography and development. The need of cooperation network in Swahili language technology”).

Takie zagadnienia jak problem dostępności zasobów, ich wielokrotnego użycia, różnych możliwości ich wykorzystania i rola licencji w przetrwaniu i dalszym rozwoju danych są poruszone m.in. w artykule „The Open-Content Text Corpus Project”.

## 5.4 Elektroniczne korpusy tekstów

Od czasów studenckich związana jestem z wieloma projektami, które miały na celu

---

1 <http://kevindonnolly.org.uk/2010/04/a-swahili-verb-analyser/> [04.05.2018]

stworzenie elektronicznych korpusów tekstów dla języka polskiego. Brałam udział w tworzeniu korpusu języka polskiego przez Wydawnictwo Naukowe PWN (omówione w wystąpieniu „On collecting data for the speech corpus of Polish Language”), następnie przez wiele lat współpracowałam z Instytutem Podstaw Informatyki PAN przy tworzeniu najpierw korpusu IPI PAN (grant KBN „Anotowany korpus pisanego języka polskiego z dostępem przez Internet”) a następnie Narodowego Korpusu Języka Polskiego, który to projekt administrowałam (grant KBN „Narodowy Korpus Języka Polskiego”). W wyniku projektu opracowany został największy anotowany korpus referencyjny współczesnej polszczyzny wielkości ponad półtora miliarda słów. Obecnie koordynuję prace w projekcie CLARIN-PL „Korpus polskich tekstów parlamentarnych XX wieku”, który ma na celu udostępnienie do przeszukiwania tekstów z biuletynów sejmowych od lat 20. XX wieku po teksty współczesne.

Moje zainteresowania wiążą się oczywiście również z korpusami języków afrykańskich i problemem niewystarczającej dostępności zasobów dla tych języków. Mając doświadczenie w projektach korpusowych dla języka polskiego, byłam współinicjatorką projektu opracowania korpusu równoległego dla języków afrykańskich w oparciu o dane wolnodostępne *Open-Content Text Corpus Project* (OCTC<sup>2</sup>). Projekt został przedstawiony w dwóch artykułach „The Open-Content Text Corpus Project” i „Open-Content Text Corpus for African languages”. Główne założenia projektu to publikacja wolnodostępnych zasobów udostępnianych poprzez SourceForge.net na licencji GNU General Public License.

### **5.5 Językoznawstwo korpusowe**

Poza praktycznym aspektem budowy korpusów, dostępnością danych czy ich formatami, korpusy wykorzystuję również jako źródło do badań językoznawczych. Badania takie, na temat konceptualizacji części ciała w języku, współprezentowałam dwukrotnie na konferencjach międzynarodowych w czasie wystąpień „Language sensitivity to body parts' terms” i „Derivation and semantic autonomy: A corpus study of Polish głowa ‘head’ and its diminutive główka”. Z drugiego wystąpienia, prezentowanego na konferencji *PLM*, jest obecnie przygotowywana publikacja.

### **5.6 Wielojęzyczność a edukacja w Afryce**

Moje zainteresowania naukowe wiążą się również z kwestią wielojęzyczności, która jest powszechna na kontynencie afrykańskim. Interesuje mnie polityka językowa państw, ze szczególną uwagą poświęconą aspektowi edukacji. Jest to temat o tyle ciekawy, że polityka językowa nie jest

---

2 <http://sourceforge.net/projects/octc> [04.05.2018]

stała i ulega ciągłej ewolucji w różnych państwach. Przykład RPA został poruszony w artykule „Droga ku wielojęzyczności RPA”.

### **5.7 Nauczanie języków afrykańskich**

Kolejny obszar moich badań związany jest z nauczaniem języków afrykańskich w Polsce. Do tej pory głównie interesowała mnie rola słowników w procesie nauczania, czego dotyczyły niektóre moje wystąpienia konferencyjne, jak np. referat, z którym zostałam zaproszona na konferencję międzynarodowego projektu *SeLA* (Scientific e-Lexicography for Africa) „Dictionaries as teaching aid for African language studies” czy referat “Dydaktyczne cechy elektronicznych słowników języka suahili” wygłoszony na konferencji „Glosa do Leksykografii”. Obecnie moje zainteresowania skupiają się na aspekcie kulturowym w nauczaniu języków obcych. Metodyka nauczania języków obcych, kwestie stosowania podejścia interkulturowego oraz rozwijania kompetencji socjokulturowych na zajęciach językowych są tematem artykułu „Elementy Kulturowe w Nauczaniu Języków Afrykańskich” opublikowanym w *Rocznikach Humanistycznych*.

Kompetencja interkulturowa stanowi nowy element nauczania języków obcych. Obecnie umiejętność dostosowania swojego zachowania do kontekstu kulturowego, zdolność radzenia sobie w sytuacjach konfliktowych, nawiązywanie i budowanie pozytywnych kontaktów międzykulturowych są równie ważne w procesie komunikacji, jak sama kompetencja komunikacyjna. W mojej dalszej pracy naukowej planuję skupić się na badaniach związanych z odmiennością kulturową i jej przejawami w języku użytkowników języków bantu.

### **5.8 Udział w projektach badawczych**

Od czasu studiów doktoranckich byłam kierownikiem (2 projekty własne plus grant promotorski), administratorem i koordynatorem (3 projekty) lub wykonawcą w sumie w dziesięciu projektach krajowych (finansowanych przez KBN lub NCN) i europejskich. Polskie projekty dotyczyły takich dziedzin jak budowa korpusów dla języka polskiego, moich badań nad leksykografią suahili i budowy Słowosieci (polskiego *wordnetu*).

Projekty europejskie z dziedziny automatycznego przetwarzania danych, w których brałam udział, to *LT4eL (Language Technology for eLearning)*, *Central and South-east europeAn Resources CESAR* oraz *PARSEME (PARSing and Multi-word Expressions) Towards linguistic precision and computational efficiency in natural language processing*. W wyniku tych prac zostały opublikowane następujące artykuły, których jestem współautorką: „Towards the automatic extraction of definitions in Slavic”, „On the evaluation of Polish definition extraction grammars”, „PARSEME – PARSing and Multiword Expressions within a European multilingual network”.

Formalnie współpracuję również z projektem Kamusi Project International, który ma na celu opracowanie wielojęzycznego słownika. Od 2015 jestem zaangażowana w polską i suahilijską część prac.

Prowadzę również współpracę z prof. Sonją Bosch (University of South Africa, RPA) przy tworzeniu ontologii morfemów języków bantu (moje opracowanie danych dla języka suahili zostało m.in. wykorzystane w artykule: Taljard, E., Faas, G. i S. Bosch. 2015. Implementation of a part-of-speech ontology: Morphemic units of Bantu languages. *NJAS* 24(2): 146-168) oraz z prof. Lionelem Posthumusem (University of Johannesburg, RPA), który jest współautorem jednego z moich artykułów na temat wielojęzyczności w krajach afrykańskich. Obecnie planuję wyjazd na Uniwersytet w Johannesburgu w celu prowadzenia zajęć, na który przyznano mi stypendium programu Erasmus Plus.

### **5.9 Dydaktyka, doszkalanie i działalność popularyzująca naukę**

Moje obciążenie dydaktyczne wiąże się głównie z nauczaniem przedmiotów podstawowych, wchodzących w zakres obowiązkowych na studiach afrykanistycznych, jak język suahili, lektura tekstów suahili, literatura suahili, wstęp do afrykanistyki, proseminarium. W obciążeniu na przyszły rok akademicki przewidziane mam również seminarium magisterskie. Metody komputerowe przydatne w pracy językoznawcy mam okazję prezentować podczas wykładu metodologicznego dla Pracowni Językowej Wydziału Orientalistycznego, który skierowany jest do wszystkich studentów wydziału. Na swoich zajęciach staram się wprowadzać nowoczesne rozwiązania, jak np. nauczanie języka poprzez platformę e-learningową UW. Przez kilka lat prowadziłam również ogólnouniwersytecki kurs internetowy „Słowniki i encyklopedie w Internecie – wprowadzenie do nowoczesnej leksykografii”, opracowany wspólnie z prof. Januszem S. Bieniem oraz internetowe autorskie zajęcia „Technologia informacyjna dla Orientalistów”.

Swoje kompetencje staram się podnosić na szkoleniach i warsztatach (takich jak np. Warsztaty z zarządzania danymi badawczymi, Przystosowanie Informatyczne - szkolenie dla instruktorów, Jak napisać doskonały wniosek, Lexicom Workshop in Lexicography and Lexical Computing, Training Course on Methods and Technologies for Consolidating and Harmonising Treebank Annotation, Fasihi ya kiswahili – literatura suahili).

Regularnie biorę też udział w konferencjach międzynarodowych, na których przedstawiam wyniki swoich badań. Dają one jedyną możliwość spotkania kolegów afrykanistów i suahilistów. Udział w dużych kongresach zagranicznych nie byłby możliwy bez wsparcia finansowego, które otrzymywałam zarówno z Fundacji na rzecz Nauki Polskiej, jak i European Science Foundation. Fundacja European Science Foundation doceniła moją pracę nad formalizmami danych,

zapraszając mnie z referatem na spotkanie „Strategic workshop on research communities and research infrastructures in the Humanities”.

W swoim dorobku mam również działalność popularyzującą wiedzę o krajach afrykańskich: jestem współautorką przewodnika „Kenia, Tanzania i Zanzibar. Praktyczny Przewodnik” wydawnictwa Pascal; organizowałam też wystawę o tanzańskim malarstwie tingatinga, która odbyła się w ramach Dni Afryki organizowanych na Uniwersytecie Warszawskim.

#### **5.10 Recenzowanie publikacji, współpraca międzynarodowa; działalność organizacyjna.**

Kilkakrotnie recenzowałam artykuły do czasopism naukowych, jak *Language Resources and Evaluation* do tomu *Special Issue on African Language Technology* (czasopismo z bazy Web of Science, lista A), *Studies of the Department of African Languages and Cultures* oraz *Afryka*.

Recenzowałam również wnioski o grant dla National Research Foundation, Thuthuka Programme, RPA.

W czasopiśmie *Studies of the Department of African Languages and Cultures* jestem członkiem komitetu redakcyjnego z funkcją redaktora, jestem również odpowiedzialna za stronę internetową czasopisma.

W swoim dorobku mam również współorganizację kilku spotkań naukowych. Byłam członkiem komitetu naukowego konferencji „Body Part Terms in Linguistic Usage: A Comparative and Typological Perspective”, która odbyła się na Uniwersytecie Warszawskim w 2017 roku. Podczas „Human Language Technology Days” organizowanym w 2012 roku przez Instytut Podstaw Informatyki PAN w Warszawie pełniłam rolę sekretarza komitetu organizacyjnego. Poza tym współorganizowałam sesję naukową „Modern Lexicography” na konferencji PLM 2007 i byłam sekretarzem komitetu organizacyjnego LXIII Zjazdu PTJ w 2005 r.

Należę do kilku towarzystw naukowych: Polskie Towarzystwo Afrykanistyczne (od 2008 r.), African Association for Lexicography (od 2009 r.), Text Encoding Initiative (w 2010 r.) oraz Chama cha Ukuzaji wa Kiswahili Duniani [Towarzystwo promocji języka suahili na świecie] (od 2017 r.).

Angażuję się również w prace wydziału. Od 2014 roku pełnię rolę przedstawiciela Wydziału Orientalistycznego w Radzie Humanistyki Cyfrowej Uniwersytetu Warszawskiego, w 2012 roku byłam członkiem Komisji Rektorskiej Uniwersytetu Warszawskiego ds. Programów Kształcenia oraz członkiem Komisji ds. Programów Kształcenia Wydziału Orientalistycznego. Przez dziesięć lat (2006-2016) pełniłam funkcję członka Komisji Dydaktycznej Wydziału Orientalistycznego UW, a także pełnomocnika Dyrekcji Instytutu Orientalistycznego UW ds. Rekrutacji (2006-2007) oraz pełnomocnika Dyrekcji Instytutu Orientalistycznego UW ds. stron internetowych (2006-2008). W latach 2005-2009 byłam opiekunem studentów Zakładu Języków i Kultur Afryki Instytutu

Orientalistycznego, byłam też sekretarzem Komisji Rekrutacyjnej Instytutu Orientalistycznego UW na rok akademicki 2005/2006.

Reasumując, jestem autorką jednej monografii, współredaktorką jednej pracy zbiorowej, oraz (współ)autorką 27 artykułów naukowych (w tym jednego w druku) oraz 5 recenzji. Większość artykułów zostało opublikowanych w czasopismach uwzględnionych na listach czasopism punktowanych MNiSW (trzy w czasopismach z listy A: *International Journal of Lexicography* i *Lexikos*, jeden w czasopiśmie z listy C: *Roczniki Humanistyczne*), pozostałe w książkach oraz recenzowanych materiałach z konferencji międzynarodowych, takich jak EURALEX, PALC, LTC czy sesji tematycznych konferencji LREC i EACL. Jestem też autorką nowego słownika suahili-polskiego opublikowanego w dwóch wersjach, elektronicznej i drukowanej oraz kilku edycji elektronicznych słowników suahili-polskiego i suahili-angielskiego opublikowanych w projekcie FreeDict.

Poza tym jestem (współ)autorką ponad 30 wystąpień na konferencjach w Polsce i za granicą, w tym takich jak EURALEX International Congress, Conference of the African Association for Lexicography AFRILEX, eLex Electronic Lexicography in the 21<sup>st</sup> Century, sesja tematyczna Language Resources and Evaluation Conference LREC, sesja tematyczna Conference of the European Chapter of the Association for Computational Linguistics EACL, Conference and Members' Meeting of the Text Encoding Initiative Consortium, Poznań Linguistic Meeting, Practical Applications in Language and Computers, oraz stricte afrykanistycznych, jak World Congress of African Linguistics, African Language Teachers Association Conference, CHAUKIDU, CHAKITA, czy Colloquium Swahili.

W swoim dorobku mam również stypendia konferencyjne m.in. z European Science Foundation czy Fundacji na rzecz Nauki Polskiej, zaproszony udział z wystąpieniami na konferencje „Strategic workshop on research communities and research infrastructures in the Humanities” (European Science Foundation, Strasbourg, Francja) i “Hildesheim Conference on Electronic Dictionaries as Information Tools” (Universität Hildesheim, Niemcy) organizowaną przez międzynarodowy projekt Scientific e-Lexicography for Africa, oraz udział w dziesięciu polskich i europejskich projektach badawczych.

*Beata Mojstowicz*