

## **Od wielojęzycznego korpusu równoległego do semantycznej mikro-typologii PERFECTUM w językach słowiańskich i bałtyckich**

Na całym świecie ludzie używają około 6500 języków. Języki różnią się między sobą, ale zróżnicowanie to jest zaskakująco usystematyzowane. Języki z różnych rodzin językowych i z odległych od siebie regionów mają wiele wspólnych cech fonologicznych, strukturalnych i semantycznych. Odkrycie uniwersaliów językowych i parametrów zróżnicowania i zrozumienie ich związku z funkcjonowaniem języka w umyśle jest kluczowe we współczesnych badaniach lingwistycznych. W odróżnieniu od uniwersaliów fonologicznych i składniowych, niewiele uwagi poświęcono uniwersaliom semantycznym. W ostatnim czasie zagadnienie to staje się przedmiotem coraz liczniejszych badań. Uczni sugerują, że poszukując uniwersaliów semantycznych powinniśmy skupić się na znaczeniu morfemów funkcjonalnych. Lista takich morfemów w każdym języku jest ograniczona, w przeciwieństwie do otwartej listy morfemów leksykalnych. Co więcej, *World Atlas of Language Structures* (typologiczna baza danych gromadząca informacje na temat struktur językowych świata) pokazuje, że morfemy funkcjonalne występują w wielu językach niezależnie od ich genetycznego pokrewieństwa i odległości geograficznej. Wiele języków posiada morfemy funkcjonalne służące do wyrażania czasu przeszłego, negacji, liczby gramatycznej, pytań. Zasadne jest więc szukanie uniwersaliów semantycznych porównując znaczenie morfemów funkcjonalnych w różnych językach.

Mimo że celem lingwistyki jest odkrywanie uniwersaliów językowych, większość badań teoretycznych opiera się na pojedynczych danych. Inną perspektywę przyjmują lingwiści komputerowi, którzy tworzą korpusy językowe zawierające miliony konstrukcji. Takie podejście otwiera szerokie spektrum możliwości badawczych. Mimo, że korzystanie z korpusów okazało się owocne w typologii językowej, nadal obserwujemy niewiele prób skonsolidowania wyników badań językoznawców komputerowych i teoretycznych. W ostatnich latach, sytuacja ta zmienia się i coraz więcej naukowców wykorzystuje metody korpusowe w badaniach nad mikro-typologią semantyczną. Celem niniejszego projektu jest dołączenie do tego nurtu badawczego wykorzystując metody korpusowe w badaniach nad semantyką kategorii PERFECTUM w językach słowiańskich i bałtyckich. Według Dahla i Vlelupillai (2013) "PERFECTUM jest używane do wyrażania wydarzeń, które miały miejsce przed momentem mówienia (lub innym punktem odniesienia), ale które są w jakiś sposób nadal istotne w tym momencie". Semantyka PERFECTUM różni się w różnych językach. Celem tego projektu jest wykorzystanie innowacyjnej metodologii zwanej TRANSLATION MINING we współpracy z naszymi partnerami z Uniwersytetu w Utrechcie do usystematyzowania tego zróżnicowania poprzez porównanie oryginalnej wersji wybranej powieści angielskiej i jej przekładu na języki słowiańskie i bałtyckie. W oparciu o te badania opracujemy mikro-typologię czasu i aspektu w badanych językach oraz opracujemy innowacyjną metodę wizualizacji dużych danych wykorzystującą technikę obliczeniową zwaną SKALOWANIEM WIELOWYMIAROWYM, która generuje mapy temporalne setek konstrukcji powiązanych z kontekstami użycia. Ten sposób wizualizacji dużych danych (niełatwych do ogarnięcia przez ludzki mózg) znacznie ułatwia testowanie teorii semantycznych i typologicznych. Metoda ta została z powodzeniem wykorzystana w badaniach nad semantyką PERFECTUM w językach germańskich i romańskich. Z pewnością, grupa języków bałto-słowiańskich, stanowiąca drugą co do wielkości rodzinę językową w Europie powinna zostać uwzględniona w tym nurcie badawczym. Języki bałtyckie zachowały liczne cechy języka proto-indoeuropejskiego. Zrozumienie ich gramatyki przybliży nas do poznania genezy rodziny języków indoeuropejskich.

Nasze badania będą miały charakter synchroniczny, ale będziemy dążyć do wyjaśnienia, jak zaobserwowane różnice w semantyce PERFECTUM wiążą się z faktami dotyczącymi zmian językowych w systemach czasowo-aspektowych języków słowiańskich i bałtyckich. Mimo że teorie oparte na badaniach korpusowych dostarczają istotnych informacji, wymagają one weryfikacji za pomocą metod eksperymentalnych testujących konkretne hipotezy przy użyciu starannie dobranej materiału badawczego. Aby odpowiednio zinterpretować dane i wyciągnąć wiarygodne wnioski, zaplanujemy wywiady z rodzimymi użytkownikami badanych języków oraz eksperymenty testujące akceptowalność konstrukcji i preferencje w użyciu form w badanych kontekstach.