

Głównym celem projektu jest odpowiedź na pytanie, **ile jest leksemów – czyli jednostek słownikowych – o kształcie TO w języku polskim**. Forma *to* jest zarówno jednym z najczęstszych, jak i najbardziej kłopotliwych wyrazów w polszczyźnie. Analizy lingwistyczne dotyczące takich form dochodzą do różnych wniosków, a informacje gramatyczne podawane w słownikach niekiedy znacznie się różnią. Co więcej, forma ta pełni wiele różnych funkcji składniowych i semantycznych, a także jest uwikłana w ciekawe zjawiska językowe: funkcjonuje jako łącznik w zdaniach takich jak (1) oraz jest zaangażowana w tematyczno-rematyczną organizację zdania (por. (2)–(3)). Mowa tu o tym, że w zdaniu (2) forma *To* wskazuje reumat – informację nową, w omawianym zdaniu reprezentowaną przez wyraz akcentowany: *Janek* – a w zdaniu (3) *to* podkreśla tematyczny status wyrazu *Janek* (który może być sparafrazowany za pomocą wyrażenia *Jeśli chodzi o Janka...*).

- (1) Janek to świetny lekarz.
- (2) To Janek kupił ten prezent.
- (3) Janek to pracuje w szkole.

Zdania (1)–(3) to zaledwie część znanych w lingwistyce użyc omawianego wyrazu. Funkcjonuje on również m.in. jako zwykły zaimek wskazujący oraz spójnik. Mimo wielu analiz poświęconych formom *to* wciąż nie jest jasne, ile należy wyróżnić leksemów o kształcie TO, jakie są ich składniowe i semantyczne właściwości oraz gdzie przebiega granica między nimi. Poprzez analizę empiryczną (korpusową i ankietową), składniową oraz semantyczną – ze szczególnym uwzględnieniem zjawisk z zakresu struktury tematyczno-rematycznej – będę starał się wypełnić tę lukę badawczą. Zakładam, że gruntowne badania empiryczne omawianych form doprowadzą do reewaluacji dotychczasowych analiz, które w ogromnej mierze bazują na przykładach skonstruowanych przez lingwistów. Co więcej, eksploracyjne badania korpusowe pozwolą na dostrzeżenie wielu różnych użyc *to*, potencjalnie niedostrzeżonych we wcześniejszych pracach. Weryfikacja dotychczasowych analiz, a także nowych hipotez i uogólnień pozwoli na sformalizowanie składni różnych konstrukcji z formami *to*. Na bazie rzetelnie ustalonej składni będzie budowana analiza semantyczna biorąca pod uwagę strukturę tematyczno-rematyczną. Takie kompleksowe badania pozwolą na zaproponowanie dobrze umotywowanej klasyfikacji leksemów TO, co będzie stanowiło istotny wkład do polskiej lingwistyki oraz potencjalny punkt odniesienia dla leksykografów.

Jedną z innowacyjnych cech projektu będzie zastosowanie triangulacji metodologicznej, nowoczesnego podejścia do badań empirycznych zakładającego kombinację różnych metodologii – w przypadku lingwistyki: badań korpusowych i ankietowych. Korpusy językowe, mimo znaczącego wpływu na jakość analiz lingwistycznych, z powodu swojego ograniczonego rozmiaru nie są w stanie zagwarantować pełnego zakresu występowania danych jednostek języka. Dopiero połączenie takich badań ze starannie zaprojektowanymi badaniami ankietowymi pozwoli na ustalenie solidnej empirycznej podstawy do dalszych analiz językowych.

Oprócz rozwiązania problemu liczby, zakresu oraz składniowych i semantycznych właściwości różnych leksemów o kształcie TO projekt będzie starał się wnieść wkład do ogólnojęzykoznawczych badań nad dwoma istotnymi zagadnieniami. Pierwsze z nich dotyczy typów konstrukcji łącznikowych występujących w językach naturalnych. Choć badania składniowe nad takimi konstrukcjami z *to* były już przeprowadzone, to jednak wciąż brakuje analizy skupiającej się na ich semantyce. Wykorzystanie nowoczesnych formalnosemantycznych narzędzi i sposobów reprezentacji do analizy konstrukcji łącznikowych z *to* może zaowocować ciekawymi wnioskami teoretycznymi. Drugie zagadnienie istotne dla lingwistyki ogólnej dotyczy możliwości formalnosemantycznego modelowania zjawisk związanych ze strukturą tematyczno-rematyczną. Konstrukcje z *to*, takie jak (2)–(3), są obiecującym obszarem badawczym, który może pozwolić na sformułowanie nowych uogólnień istotnych dla dziedziny.

Warto dodać, że wynikiem empirycznej części projektu będzie baza stwierdzeń teoretycznych dotyczących statusu gramatycznego i liczby leksemów o kształcie TO zawartych w dotychczasowej literaturze naukowej, wzbogacona o informację, czy dany sąd znajduje pokrycie w badaniach ankietowych i korpusowych. Baza ta ma stanowić istotny zasób wiedzy dla badaczy języka polskiego. Wyniki badań składniowych i semantycznych będą zgłoszone w formie artykułów naukowych do renomowanych polskich i zagranicznych czasopism lingwistycznych (ogólnych, slawistycznych i semantycznych) oraz zostaną zaprezentowane na polskich i zagranicznych konferencjach naukowych.