

## Streszczenie

Tematem pracy jest zjawisko faktywności we współczesnej polszczyźnie. Rozważone zostały przede wszystkim następujące trzy kwestie:

1. Które czasowniki w języku polskim są faktywne, a które takie nie są?
2. Jak często i w jaki sposób użytkownicy języka polskiego posługują się czasownikami faktywnymi?
3. W jaki sposób działają modele komputerowe przewidujące, czy w danym wypowiedzeniu zachodzi relacja wynikania, sprzeczności bądź niezależności między całym wypowiedzeniem a jego dopełnieniem i jakie dane językowe są potrzebne, by podnosić ich skuteczność?

By odpowiedzieć na pierwsze pytanie, przeanalizowano 325 prymarnie propozycjonalnych czasowników współczesnej polszczyzny. Każdy czasownik opatrzony został przykładem użycia i informacją o tym, czy jest faktywny.

Odpowiedź na drugie pytanie wynika z analizy odpowiednio wybranych 2348 wypowiedzeń z Polskiego Korpusu Koreferencyjnego (PKK), który stanowi reprezentatywną próbkę Narodowego Korpusu Języka Polskiego (NKJP). Uwzględnione zostały wszystkie wypowiedzenia o strukturze „ $V\{ze\}p$ ”, które znajdują się w PKK. Przykładem wypowiedzenia o takiej strukturze jest (1):

(1) *Teraz już wiem, że był to zapach cebuli.* [NKJP]

To, co mnie w takich wypowiedzeniach szczególnie interesowało, to: (a) rodzaj predykatu głównego, wokół którego osnute jest całe zdanie (faktywny / niefaktywny) i (b) relacja między całym zdaniem a będącym jego częścią dopełnieniem zdaniowym (wynikanie, sprzeczność i niezależność).

Ustalenie odpowiedzi na ostatnie, trzecie pytanie wymagało zbudowania odpowiednich modeli komputerowych. W ramach przeprowadzonych prac na Wydziale Matematyki i Nauk Informacyjny Politechniki Warszawskiej powstało kilka takich modeli. Ich implementacji, czyli wytrenowaniu i przetestowaniu posłużył przygotowany przeze mnie materiał językowy. Następnie, za pomocą odpowiednich narzędzi,

przeanalizowałem ich sposób działania, skupiając się przede wszystkim na ustaleniu elementów, które są konieczne do poprawy ich skuteczności.

Rozprawa składa się z pięciu rozdziałów. W pierwszym z nich skupiam się na zaprezentowaniu zintegrowanej siatki pojęciowo-terminologicznej, która znalazła zastosowanie w pracy. Siatka ta w zamierzeniu ma być łatwo przekładalna na alternatywne koncepcje pojęciowe, a jej ewentualny przekład powinien być irrelevantny dla uzyskanych wyników badań.

Drugi rozdział rozpoczynam od przedstawienia kryteriów, które stosuję do weryfikacji przeciwstawnych wobec siebie hipotez na temat faktywności danych jednostek leksykalnych. Po zaprezentowaniu i omówieniu sposobu, w jaki badam, czy dana jednostka jest faktywna, przechodzę do omówienia wybranych wyrażen. Analizowane czasowniki podzieliłem na następujące cztery klasy semantyczne: czasowniki *epistemiczne*, *percepcyjne*, *mówienia* i *emotywne*. Zdecydowanie najwięcej uwagi poświęcam epistemicznemu czasownikowi *wiedzieć, że\_*, który jest kluczową jednostką w klasie wyrażen faktywnych.

Trzeci rozdział zawiera przede wszystkim liczbowe wyniki analiz wypowiedzeń z PKK. W rozdziale zaprezentowane są między innymi takie dane, jak np. stosunek czasowników faktywnych do niefaktywnych czy współwystępowanie takich cech jak niefaktywność i wynikanie.

W czwartym rozdziale skupiam się na wynikach wytrenowanych za pomocą uczenia maszynowego modeli i na analizie ich działania. Przyglądam się modelom, które trenowane były zarówno na podstawie cech w stworzonym przeze mnie zbiorze danych (np. klasa semantyczna czasownika, jego czas gramatyczny, obecność negacji wewnętrznej), jak i wyłącznie całego wypowiedzenia (tzw. *word embedding*). Te ostatnie bazują na modelu HerBERT, który jest polskojęzyczną wersją modelu BERT. Głównym zadaniem wytrenowanych modeli było przewidywanie jednej z trzech relacji: wynikania, sprzeczności, niezależności, między całym wypowiedzeniem a stanowiącym jego część dopełnieniem. Porównuję ze sobą wyniki poszczególnych modeli, z uwzględnieniem wyników uzyskanych dla trzech powyższych relacji logicznych.

Rozdział piąty to podsumowanie całej pracy. Wracam w nim do postawionych pytań badawczych i w skrótovej formie formułuję i omawiam uzyskane odpowiedzi.