

Kwantyfikatory w języku: użycie i znaczenie

Język naturalny jest głównym narzędziem ludzkiej komunikacji, zarówno w nauce jak i życiu codziennym. Język jest szeroko badany z różnych teoretycznych perspektyw: językoznawczej, logicznej czy obliczeniowej. W centrum tych interdyscyplinarnych badań leży prosta lecz bardzo wpływowa koncepcja znaczenia, która okazała się niezwykle istotna dla wielu zastosowań językoznawstwa, np. w informatyce przy tworzeniu teoretycznych podstaw dla systemów dialogowych, automatycznego rozumowania, czy internetowych funkcji wyszukiwania. Ciągle pozostaje jednak pytanie w jakim stopniu ta koncepcja może wytłumaczyć naturalne ludzkie zachowania językowe. Ostatnie lata były świadkami rosnącej współpracy między językoznawstwem i naukami o poznaniu. Narodziła się nowa dyscyplina eksperymentalnej semantyki i pragmatyki. Jedną z głównych trudności w tej dyscyplinie jest brak naturalnych danych językowych. Badania koncentrują się zatem na pozyskiwaniu danych przez eksperymenty psycholingwistyczne. Tego typu dane są podstawowe dla zrozumienia jak ludzie przetwarzają język naturalny, lecz niestety są one również bardzo problematyczne: Badani nie zawsze podążają za swoją naturalną intuicją językową, często zamiast tego rozwijają wyspecjalizowane strategie dostosowane do specyficznego zadania eksperymentalnego. Można powiedzieć zatem, że takie dane nie są lingwistycznie wiarygodne. Celem tego projektu jest przezwycięzenie tej zasadniczej trudności i dostarczenie lingwistycznych danych opartych na naturalnym użyciu języka. W projekcie zbudujemy i przeanalizujemy polski korpus, czyli gigantyczną bazę danych zawierających zdania języka polskiego specjalnie anotowane informacją lingwistyczną. W szczególności skupimy się na językowych wyrażeniach, które mówią o ilości, tzw. kwantyfikatorach, jak np. `niektóre` czy `większość`. Korpus zostanie opublikowany online w wolnym dostępie wraz z narzędziami do automatycznego wykrywania wyrażen kwantyfikatorowych. Opierając się na analizie zbudowanego korpusu będziemy również w stanie lepiej zrozumieć dlaczego ludzie używają pewne wyrażenia kwantyfikatorowe znacznie częściej niż inne.