

Ramy semantyczne w tekstach prawnych – analiza konfrontatywna polsko-angielsko-niemiecka

Agnieszka Pluwak

Cele badawcze pracy

Analiza leksykalna słownictwa związanego z pojęciem *najmu*, niezbadanego dotąd metodą ramową, a zatem:

1. wkład w budowę FrameNetu dla języka polskiego; pośrednio – przyczynienie się do rozwoju zasobów leksykalnych, stosowanych w przetwarzaniu języka naturalnego;
2. przygotowanie materiału językoznawczego (leksyka, elementy ramy, konstrukcje) do zakodowania, czyli przełożenia na język reguł informatycznych, celem stworzenia programu do automatycznego wyszukiwania informacji w tekstach umów prawnych.

Tezy:

Główne tezy pracy są ujęte w jej poszczególnych rozdziałach:

1. *Teorię semantyki ramowej w ujęciu Fillmore'a da się zastosować do analizy wielu języków, gdyż postulowane przez niego ramy mają charakter pojęciowy (semantyczny).*
2. *Zastosowanie semantyki ramowej jako metody do analizy języka polskiego:*
 - A. *rodzi dylematy natury formalnej - w dopasowaniu do istniejących już ram ze względu na cechy charakterystyczne dla budowy języka słowiańskiego takie, jak aspekt czasownika, rozbudowane prefiksy, podmiot domyślny zawarty w czasowniku - ale nie rodzi problemów natury pojęciowej ze względu na ogólny charakter definicyjny ram;*
 - B. *stwarza konieczność zbudowania nowych ram dla niektórych nieopisanych dotychczas we FrameNecie znaczeń;*
 - C. *daje możliwość zastosowania wyników analizy w tworzeniu systemów NLP¹.*
3. *Semantyka ramowa w ujęciu FrameNet znajduje zastosowanie w wielu funkcjach przetwarzania języka naturalnego:*
 - A. *jako projekt leksykalny (oznaczony przez lingwistów korpus tekstów prasowych służący do trenowania systemów uczenia maszynowego),*
 - B. *poszczególne opisane ramy znajdują zastosowanie w funkcjach ekstrakcji danych typu oznaczanie ról semantycznych (ang. semantic role labeling) jako schematy będące szkieletem całego systemu NLP.*
4. *W badaniach językoznawczych (zarówno językoznawstwa tradycyjnego jak i komputerowego) istniało wiele podejść do analizy argumentów (walencja syntaktyczna i semantyczna) i jej zastosowania. Jednak w przypadku analiz dziedzinowych takich, jak język prawniczy*

¹ NLP (ang. Natural Language Processing) – przetwarzanie języka naturalnego, zwane też językoznawstwem komputerowym.

opisujący przypadek najmu, potrzebny jest opis walencji semantycznej (a nie jedynie syntaktycznej) na bardziej szczegółowym poziomie (niż VerbNetu czy PropBanku) i dlatego wybrano podejście semantyki ramowej w ujęciu FrameNet do analizy wybranego materiału.

Powyższe tezy te składają się na tezę główną, która brzmi:

Możliwym jest wpisanie leksemów z różnych języków, w tym – dla języka polskiego - w ramy semantyczne zdefiniowane w projekcie FrameNet lub zbudowanie nowych ram semantycznych dla słownictwa prawniczego, pojawiającego się w umowach najmu, na ich podstawie – zdefiniowanie konstrukcji semantyczno-gramatycznych, którymi wyrażane są poszczególne elementy ramy, a następnie przełożenie ich na reguły systemu informatycznego.

Materiały badawcze

Do analizy wybrano słownictwo prawnicze (około 20-30 jednostek leksykalnych, głównie czasowników, ale też rzeczowników, przymiotników i przysłówków) z dziedziny najmu nieruchomości. O doborze materiałów zdecydowały następujące czynniki:

- A. fakt, iż słownictwo to nie zostało dotychczas opisane w projekcie RAMKI, co umożliwia poszerzenie zasobów projektu, mającego szansę stać się początkiem polskiego FrameNetu;
- B. dostępność materiałów w ww. językach, to znaczy prawnych umów najmu,
- C. możliwość praktycznego wykorzystania projektu do automatyzacji analizy umów najmu w jednej z największych firm z branży zarządzania nieruchomościami w Europie.

Metody badawcze

Semantyka ramowa w ujęciu FrameNet jest podejściem szeroko stosowanym w ekstrakcji danych z tekstów, gdyż jest to projekt leksykograficzny, który od podstaw zakładał wykorzystanie swoich zasobów w lingwistyce komputerowej. Analiza w ujęciu FrameNet tym różni się od innych podejść do analizy argumentów, że:

1. jest to analiza głównie semantyczna na poziomie szczegółowym (jest to opis na poziomie kategorii kognitywnych, w którym przypisuje się elementom ramy konkretne role sytuacyjne, a nie na bardziej ogólnym poziomie agenta, jak ma to miejsce w innych projektach leksykalnych np. VerbNet);
2. jest to analiza również syntaktyczna, która pokazuje, jak elementy składniowe realizują poszczególne role semantyczne;
3. jest to zapis umożliwiający zebranie jednostek leksykalnych w zbiory o wspólnej części znaczenia.

Metodę ramową wybrano do analizy materiału, gdyż realizacja celów pracy wymaga podejścia bardziej szczegółowego niż w innych projektach leksykograficznych, a podejście FrameNet zawiera zarówno semantykę, jak i konstrukcje syntaktyczne, umożliwiające przekład na system reguł do przetwarzania informatycznego.

