

СТАНОВИЩЕ

за дисертацията за придобиване на образователната и научната степен *доктор* на Ивелина Стоянова на тема „Автоматично разпознаване и тагиране на съставни лексикални единици в българския език”

от проф. дфн Руселина Ницолова

Темата на дисертацията на Ивелина Стоянова „Автоматично разпознаване и тагиране на съставни лексикални единици в българския език” е много сполучливо избрана и в квантитативен, и в квалитативен аспект. Лексикалните единици, съставени от две или повече графични думи, наречени *несвободни фрази*, представляват 24, 49% от единиците в българската семантична мрежа БулНет, като има редица нерешени проблеми при автоматичното им разпознаване и аотиране в компютърния анализ. Целта на работата е изграждане на теоретичен модел и неговото практическо приложение за автоматично разпознаване на съставните лексикални единици. И. Стоянова поставя много ясно основните задачи на дисертацията: да бъде дефиниран точно обектът на изследване, като бъдат анализирани трудните гранични случаи; да бъде създаден теоретичен модел, който да описва СЛЕ на морфо-синтактично, семантично, синтактично и прагматично ниво; да се изработи детайлна класификация, която да има практическо приложение за идентификацията, описанието и анализа на обекта; да се разработи методология за разпознаване на СЛЕ и отделните им категории; методологията да бъде приложена в практиката, като бъдат анализирани резултатите и възможностите за усъвършенстване.

Интересни са принципите на изследването: необвързаност с конкретна школа или теория, но използване на досегашния опит в различните области и течения; обзорност, последователност, практически характер и ограничаване на обекта за изследване до именните фрази, за да стане задачата изпълнима. Материалът е събран от Уикипедия, като се използват и лексикални ресурси от Секцията за компютърна лингвистика, специално създадени ресурси и материал от тълковния речник на Попов.

Дисертацията с обем 223 стр. съдържа 6 глави, от които първата е увод, библиография и две приложения: за предварителната обработка и списък на

ресурсите. В увода са представени целите, задачите и методологията, използваните лексикални ресурси и структурата на дисертацията. Във втората глава е предложен много ясен преглед на литературата и терминологията, като са разгледани най-важните англоезични автори, Мелчук и българските автори. И. Стоянова предлага и дефиниция на СЛЕ, като ги съпоставя със сложните думи, колокациите, наименованията и терминологията. Тя приема и класификацията на несвободните фрази на Болдуин и др. (2003), според която те се делят на неразложими, идиосинкретично разложими и прости разложими (СЛЕ), като с пълно основание добавя в класификацията и съставните лексикални единици със служебна функция (съставни съюзи и предлози) и свободните колокации, които са само статистически маркирани, без да са маркирани по другите езикови показатели. Дисертантката дава и две дефиниции на СЛЕ: теоретична и с практическа насоченост. Тя представя композиционалността като основна характеристика на СЛЕ.

Третата глава съдържа лингвистично описание на СЛЕ, което разглежда семантичните, синтактичните и прагматичните им особености. Авторката предлага и редица убедителни аргументи срещу включването на СЛЕ в речника. Тя прави и приложна класификация според характеристиките на идиоматичността, която се изразява в степента на конвенционалност, разбираемост и композиционалност.

Четвъртата глава е посветена на автоматичното разпознаване и тагиране на СЛЕ в българския език. И. Стоянова прави много задълбочен анализ на трите основни метода за разпознаване на СЛЕ: лингвистични методи, разчитащи на лингвистичен анализ и ресурси, количествени методи – основани на количествена информация за единиците и статистически тестове и хибридни методи, съчетаващи първите два метода. Дисертантката прави и анализ на факторите, които влияят върху ефективността на методите. Тя разработва и компютърна програма за разпознаване и тагиране на СЛЕ и се спира на методите за оценка на резултатите: метод въз основа на честотен анализ и филтриране на синтактични конструкции, метод, основан на лингвистични правила, при който се постига завидната точност от 91.51%.

Петата глава представя точно приносите от изследването и са набелязани насоки и идеи за бъдеща работа. Библиографията на работата е много голяма - 16 стр. и доста пълно представя различните автори, свързани с разглежданата проблематика.

В заключение би могло да се обобщи, че разглежданата дисертация е труд с несъмнен приносен характер в областта на компютърната лингвистика. Авторката представя за пръв път СЛЕ в теоретичен и компютърнолингвистичен аспект. Тя прави теоретичен анализ на своя обект на семантично, граматично и прагматично равнище, като въвежда и собствена добре дефинирана терминология. Предложена е класификация на СЛЕ с подчертано приложен характер въз основа на тяхната идиоматичност, както и относително проста методология за разпознаване на единиците на етапи несвободни фрази, СЛЕ, отделни подвидове според различните класификации. Отчетена и ефективността на предложената методология. Освен това е разработена и система от компютърни модули, които могат да бъдат приложени за различни задачи. Дадени са и някои насоки и идеи за бъдеща работа. Езикът на дисертацията е много ясен, термините са дефинирани, като авторката много добре определя мястото на собствените си възгледи на фона на световното развитие на науката. Дисертацията представя И. Стоянова не като начеващ, а като вече зрял изследовател с отлично конципиран и изработен труд, който е само един етап в творческото ѝ развитие, очертаващо се да бъде много успешно.

Всичко това ми дава достатъчно основание да предложа на почитаемото Научно жури да присъди образователната и научната степен *доктор* на Ивелина Стоянова за дисертацията ѝ „Автоматично разпознаване и тагиране на съставни лексикални единици в българския език”.

София, 27.05.2012 г.

Подпис:

(проф. дфн Руселина Ницолова)