

## Становище

за дисертацията на Ивелина Стоянова „Автоматично разпознаване и тагиране на съставни лексикални единици в българския език” за присъждане на образователната и научна степен „доктор”  
от проф. дфн Лилия Крумова - Цветкова

Предложеното за защита изследване „Автоматично разпознаване и тагиране на съставни лексикални единици в българския език” с автор Ивелина Стоянова е особено актуално – то отговаря на потребности на теорията и практиката на компютърната лингвистика. Лексикалните единици, състоящи се от две или повече графични думи, наречени несвободни фрази, не са достатъчно изследвани, макар че са проблем на синтаксиса, лексикологията и фразеологията. В случая погледът към тях е от позициите на компютърната лингвистика. Не е отделяно достатъчно внимание на една специфична част от несвободните фрази – съставните лексикални единици. Тяхното идентифициране, проучване и разграничаване от една страна от свободните фрази, а от друга – от устойчивите фрази, е от съществено значение за автоматичната обработка на естествения език. Асиметрията между форма и съдържание при тях – формално са разделнооформена единица, а семантично – цялостна – създава затруднение при компютърната обработка на езика. Всичко това определя и дисертабилността на темата.

Целта и задачите са добре формулирани: Като се очертаят основните теоретични и практически проблеми на съставните лексикални единици, да се изгради теоретичен модел и се разработи методология за практическото приложение на този модел за целите на автоматичното разпознаване и тагиране на съставните лексикални единици, като част от лингвистичния анализ и анотация на български текстове. Това определя и добре структурираното изложение на труда.

В работата си докторантката се опира и строго се придържа към ясни и точни принципи: независимост, обзорност, последователност, практическа насоченост, изпълнимост..

Изследването се базира на изключително богати езикови ресурси, което също така е фактор за успешност.

Докторантката е дала решение на поставените пред изследването задачи:

1. Като се опира на комплексния модел за описание на типовете маркираност на лексикалните единици и на класификацията на несвободните фрази на Болдуин, авторката представя своя класификация на несвободните фрази, като определя мястото на свободните лексикални единици сред тях.
2. Дадени са сполучливи дефиниции – теоретична и практическа – на термина *съставна лексикална единица*. Очертани са границите на понятието, като са намерени критерии за разграничаването на съставните лексикални единици от свободните колокации и от другите типове несвободни фрази. Практическата дефиниция е от значение за изграждане на методология за компютърно описание на СЛЕ.
3. Описана е характеристиката на съставните лексикални единици. Анализирани са стегнатото, но последователно и убедително семантичните, синтактичните, прагматичните и формообразуващите особености на СЛЕ.

4. Разгледано е отношението на СЛЕ с близки категории като сложни думи, наименования, терминология. Това се налага, поради частичното застъпване на тези категории и възникващите на тази основа проблеми при обработката им.
5. Представени са класификации на СЛЕ въз основа на семантични и синтактични признаци, както и класификация според признака идиоматичност. Последната класификация е определена като приложна, тъй като в нея се предлагат формални показатели за целите на автоматичната идентификация на СЛЕ. Както се разбира, признакът идиоматичност съвместява три признака: конвенционалност, разбираемост и композиционалност, което означава, че се отнася до форма, съдържание и фреквентност на СЛЕ. Несъмнено тази класификация, насочена към решаване на практически задачи, е приносна.
6. След представяне и оценка на основните методи, използвани за автоматичното разпознаване на несвободните фрази и СЛЕ и преглед на компютърните програми, прилагани за тази цел, е изложена разработена от докторантката компютърна програма за експериментално приложение на разработените методи за разпознаване и тагиране на СЛЕ.

Програмата се занимава с цялостна обработка и анализ на корпуси. За разпознаването на СЛЕ се предлага определен модел, включващ алгоритъм, основан на определени лингвистични правила. Резултатите се оценяват с честотна характеристика. Програмата се състои от отделни модули, приложими за отделни задачи. Усъвършенстването ѝ и добавяването на нови функционалности, по думите на авторката, е бъдеща задача.

Приносен момент в работата е, че в нея се поставят задачи за бъдещи изследвания, насочени към цялостно разработване на темата.

Смятам, че изследването на Ивелина Стоянова с конкретно поставените и разрешени проблеми на СЛЕ е приносно за българската лексикология и компютърна лексикография.

Имам обаче някои забележки относно терминологичния апарат на изследването. В българската лексикология термини като *словосъчетание*, *фразеологизъм*, *идиом*, *идиоматичност*, *фразема* са утвърдени и имат установено съдържание, изградено под влияние предимно на руското езикознание и руската фразеологична школа. В настоящото изследване се привеждат нови за българското езикознание терминология и класификации на английскоговорящи автори. Това е положителен факт, но мисля, че между традиция и настояще трябва да има някаква връзка. Изобщо много скромно място е отделено на изследванията в българското езикознание при представяне на основните възгледи за несвободните фрази и класификацията им. Вярно е, че в приложен аспект в българското езикознание не е работено достатъчно, но в теоретичен план приносите съвсем не са малко.

Не приемам термина *съставна дума*, защото противоречи на основното условие в дефиницията на думата – че е цялостнооформена единица.

В изложението и в Библиографията е цитиран неточно многотомният академичен Речник на българския език, издаван от Института за български език – посочен е като Речник на съвременния български език.

Свободните лексикални единици са част от словното богатство на езика, което изисква специално внимание при превод, за да има пълна еквивалентност – семантична и прагматична, при представяне на съответните понятия. Ето защо смятам, че присъствието на СЛЕ в речниците – традиционни или електронни, е необходимо. Всъщност самата

авторка посочва , че за наименованията трябва да се търси не само валиден, но и установен превод. Това се отнася и за терминологията, за поздравите и под. Това определя и необходимостта от лексикографско представяне на тези категории имена.

Въпреки тези критични бележки, впечатлението ми от предложеното изследване и от стила и работата на Ивелина Стоянова е положително. Изследването показва, че докторантката има качествата на изследовател с умение да анализира, да вижда, поставя и систематизира проблемите, да се изразява стегнато, точно и ясно. Стилът и езикът са на необходимата висота.

Успешното решаване на поставените задачи пред изследването ми дава основание да предложа да бъде присъдена научната и образователна степен „доктор” на докторантката Ивелина Стоянова.