АВТОРСКА СПРАВКА
на Владимир Борисов Периклиев

По чл.29, т.3 от ЗРАСРБ, съотв. чл. 60 т.3 от ПЗРСАРБ - научни публикации:

Общият брой на научните публикации е 71, от които 52 са самостоятелни, а 49 са публикувани в чужбина. Те включват 2 публикувани монографии в чужбина и една приета за печат (вж. прил. **Други докум1.pdf**), 35 статии в списания (31 от които включени в European Reference Index for the Humanities (ERIH 2011) и/или с Thomson Reuters импакт фактор (IF 2010)) и 33 доклада/резюмета (28 от които в утвърдени международни конференции/конгреси). 3 статии са написани по покана във водещи международни списания (с ERIH INT1 или IF), 1 статия в некласифицирано международно списание, а 2 доклада в утвърдени международни конференции/конгреси (вж. прил. **Публикации общо.pdf**).

Представените за конкурса научни публикации са 32, от които 22 са самостоятелни, като всички са публикувани в чужбина. 5 тях [17, 18, 28, 29, 30] са представяни за получаване на научната степен „доктор" или академичната длъжност „доцент". Останалите 27 работи са публикувани след това – спазено е изискването на чл.2 т.6 от **Правилника на ИМИ за приложение на ЗРАСРБ**. Представените публикации включват: 2 монографии, 18 статии в международни списания (14 от които с ERIH (2011), категории INT1 или INT2, и/или импакт фактор), 12 доклада в утвърдени международни конференции/конгреси. 3 статии са написани по покана във водещи международни списания (с ERIH INT1 или IF), а 1 статия в некласифицирано международно списание (вж. прил. **Публикации за конкурса.pdf**).

Тематично представените работи попадат в следните три направления:

А. Машинна обработка на естествен език (8): 16, 18, 24, 26-30

В рамките на това направление се обособяват три основни тематики:

А1. Неоднозначност
Неоднозначността (ambiguity) е основен проблем пред системите за разбиране и машинен превод, което предполага както нейното изучаване (в отделни езици и сравнително), така и предлагане на ефективни методи за нейното разрешаване.

В статия [18] (публикувана в официалния орган на Американската асоциация по славистика) се изолират различни типове синтактична неоднозначност в рамките на „граматиката на зависимостите" (dependency grammar) и се описва подробно един от тях в български. Важен резултат е, че всички логически възможни конструкции от изучавания тип, се оказват и реално възможни в български.

В [30] се предлага следният метод за разрешаване на синтактична неоднозначност в машинния превод: на конструкцията от езика, от който се превежда, се съпоставя неоднозначна конструкция в езика, на който се превежда. Този нов подход се базира на подробното емпирично съпоставително изучаване на английски и български и от общи съображение е ясно, че той е валиден и за много други двойки (напр. индо-европейски) езици. Докладът е изнесен на COLING84 в Станфорд и този резултат е отбелязан от редица авторитетни специалисти в машинния превод (вж. прил. **Цитати.pdf**).

Неоднозначността се разглежда и в други работи извън представените за конкурса (вж. прил. **Публикации общо.pdf**; напр. [30] от списъка на всички публикации е дала повод за написване на статия от редактора на сп. „Български език" и двете излизат под общото заглавие „Въпроси на езиковата неоднозначност", вж. прил. **Цитати.pdf**; [63] и [67] са доклади в същата тематика съответно на 28-ми Световен конгрес по приложна лингвистика, Сидни 1987, и 7-ми Международен конгрес по логика, методология и философия на науката, Залсбург 1983.)

А2. Словоред

Словоредът е друга важна тематика в системите за обработка на езика. За разлика от английски език, който има фиксиран словоред, повечето други езици, и в частност български, имат различни степени на свобода на наредбата, което изисква създаването на подходящи формализми за тяхното изразяване.

В съвместния доклад [29] се прави автоматизирана проверка на словоредната хипотеза за „проективността", дефинирана в рамките на граматиката на зависимостите. Хипотезата се проверява за български език, като се показват редица контра-примери. В същото време се аргументира тезата, че въпреки тези контра-примери, това словоредно ограничение е важно в системите за обработка на българския език.

Съвместният доклад [27] предлага формализма FOG (Flexible word Order Grammar) и програмната му реализация, който е средство за удобно изразяване на граматики със свободен словоред в термините на т. нар. Definite Clause Grammar (DCG), като се описват предимствата на разширения DCG-формализъм, отнасящи се до неговата икономичност на изразяване на сложни правила, модулност и декларативност.

Друга съвместна работа [26] е продължение на горната и предлага формализма EFOG (Extended Flexible word Order Grammar) и има сходната цел за разширяване на експресивната сила на граматиката, но се отнася до коренно различен тип граматика от DCG, а именно до форма́та Immediate Dominance/Linear Precedence на Generalised Phrase-Structure Grammar.

В [24] се предлага система, която машинно се обучава от примери на словоредните правила (Linear Precedence) на граматика от типа Immediate Dominance/Linear Precedence.

Проблемите на словореда се разглеждат и в други работи извън представените за конкурса (вж. прил. **Публикации общо.pdf**; напр. [46] от списъка на всички публикации е поканен доклад на панелна дискусия на 16-ти Международен конгрес на лингвистите в Париж през 1997 г.)

А3. Генерация на текст

В съвместната работа [28] се описва експериментална система за обучение по математика. Системата приема на входа ограничен клас от алгебрични задачи с думи, формулирани в свободен текст, и може да реши тези задачи, както и да предложи една или повече „подсказващи" перифрази на тези задачи, които облекчават тяхната формализация, т.е. превеждането им в уравнения.

Работа [16] разглежда проблемите на анализ и генерация в рамките на т. нар. „Референтна граматика" (Reference Grammar), предложена от именития шведски лингвист Бенгт Сигурд, като се фокусира върху проблемите на именната фраза в български език и решаването им в този модел.

В също [21].

В. Машинно откритие в лингвистиката (17): М-2, 1, 4, 8-15, 17, 20-23, 25

Машинното откритие е приложна област на изкуствения интелект, чиято задача е да изолира важни научни проблеми и да създава и прилага компютърни програми за тяхното решаване. Основополагащите изследвания в областта са проведени в Университета „Карнеги Мелън" под ръководството на Хърбърт Саймън, където авторът е поканен да работи 6 месеца с Раул Валдес-Перез, бивш докторант и сътрудник на Саймън. Това сътрудничество продължава повече от 4 години. Работата на автора по тази проблематика датира от около 20 години.

В статия [17] се разглежда използването на евристики в лингвистиката. В единствената предишна лингвистична работа, в която се споменава понятието „евристика", то се споменава само накратко и то в контекста на философията на науката, без ясни препратки към лингвистиката. Работата за първи път по-подробно разглежда различните евристики на Поя и ги илюстрира, решавайки два значими лингвистични проблема, единият от които представлява опровергаване на широко приемана теза на Н. Чомски. Подобна тематика се разглежда и в други работи извън представените за конкурса (вж. прил. **Публикации общо.pdf**; напр. [60] от списъка на всички публикации докладва този резултат на 14-тия Международен конгрес на лингвистите в Берлин през 1987 г.; срв. също [53] на 15-тия Международен конгрес на лингвистите в Квебек през 1992 г. и [62] на 8-ми Международен конгрес по логика, методология и философия на науката в Москва през 1987).

В доклада [25] се разглеждат проблемите на емпиричното откритие в лингвистиката. Особено внимание се отделя на критиките на Н. Чомски на т. нар. „процедури на откритие" (discovery procedures) и вредното влияние, което те са оказали на лингвистиката в исторически план. Описва се система, моделираща каноните на Джон Стюарт Мил, и кратко се разглеждат няколко от откритията на системата (напр. известния закон на Грим). Работата е докладвана на Пролетния симпозиум на Американската асоциация по изкуствен интелект (с участници двама Нобелови лауреати, вкл. Хърбърт Саймън, и един член на Американската академия на науките) и представляваше единственият доклад от хуманитарната област.

[15] е статия по покана от списанието *The Knowledge Engineering Review* (Cambridge) за коментар върху статия, отнасяща се до машинното научно откритие, на Раул Валдес-Перез (председател на програмния комитет на конференцията, спомената в предишния параграф).

[23] е съвместна работа с Валдес-Перез, която описва начален вариант на системата MPD (Maximally Parsimonious Discrimination). При зададени класове и примери на тези класове, описани в термините на атрибут-стойност, тази система открива максимално икономичното им разграничение, използвайки минимален брой глобални атрибути и най-кратки описания на класовете.

Забележка. Тази и следващите съвместни работи с Валдес-Перез са проведени в рамките на финансирания от US NSF проект "Generic Tasks of Scientific Discovery". При описаните в публикациите програми всеки от нас поддържаше отделна компютърна реализация, съответно на Лисп и Пролог.

[22] е съвместна работа с Раул Валдес-Перез, която описва накратко различни възможни лингвистични приложения на програмата MPD. Задачата за икономично разграничаване на класове, моделирана от MPD, е обща по характер и следователно е потенциално използваема за решаването на различни задачи от различни области на лингвистиката (а и извън нея). Основната трудност е в изолирането на подобни

автоматизируеми задачи. В доклада се показва приложимостта на метода в различни задачи в областите на лексикологията, фонологията, типологията и речевата патология.

Съвместната статия [14] е първата работа автоматизирала т. нар. „компонентен анализ" на термините за родство, тематика владяла лингвистиката и социалната антропология за десетилетия и продължаваща да има своето влияние и днес. Разполагайки с роднинските термини на един език и техните „описания" (напр. в бълг. *чичо* = брат на баща ∨ брат на майка), програмата KINSHIP генерира техния „смисъл" (*чичо* = мъжки пол & +1 поколение & родство по неправа линия), така че всеки термин се разграничава от всички останали по поне един признак, като при това анализът е максимално икономичен (използва минимален брой глобални признаци и най-кратки дефиниции на термините). KINSHIP е открила компонентните анализи на няколко езика, като този за английски – който е разглеждан в редица работи – се оказва не само най-икономичният (критерий приет в областта), но и единственият даващ желаните конюнктивни дефиниции на всички роднински термини от английски. Статията е публикувана в най-авторитетното списание в областта *Anthropological Linguistics*.

В [9] се използват разработени от нас програми за разграничаването на индивиди и те са използвани за профилиране на 451 езика (базата UPSID от Университета на Калифорния в Лос Анджелис) в термините на техните фонологични звукове, като (машинно) са направени лингвистически интересни наблюдения и обобщения върху структурата на получените профили.

Съвместната статия [11] дава най-подробното описание на алгоритъма ни за разграничаване на множество от класове и дава илюстративни примери на приложение на програмата и извън лингвистиката (напр. в психологията и химията). Системата може да разграничава класове по най-икономичен начин (в смисъла, описан по-горе), като това разграничение може да бъде както абсолютно, така и приближено. В допълнение програмата може сама да създава „производни атрибути" чрез логически операции върху оригиналните, които да се използват в случай, че последните са недостатъчни за разграничението на всички класове. В Приложение към статията нашият метод подробно се разграничава от класическия метод C4.5 на Куинлън.

Друга застъпена тематика в представените работи представлява откритието на т. нар. „универсали" в лингвистиката, т.е. твърдения, валидни във (почти) всички езици. Универсализмът е водещо течение в лингвистиката в последните десетилетия.

Една от пионерските работи в областта е [13], която машинно открива значимите универсали от роднински модели (kin term patterns) в база от над 500 езика на един от основоположниците на съвременната социална антропология Джордж Мърдок. Работата е цитирана в сп. "Science".

Работа [21] описва UNIVAUTO (UNIVersalsAUthoringTOol), първата програма за научно откритие, която е в състояние да представи своите открития във формата на научна статия. Системата разполага на входа с информация за представително множество от езици, представена като обект-атрибут-стойност, (евентуално) откритията на агент-човек, направени върху същите данни, както и известна друга информация. Модулът за откритие UNIV открива универсали (абсолютни или статистически) от различен логически вид (типично, дву- или повече съставни импликации) и проверява тяхната статистическа значимост (с пермутационен тест или χ-квадрат). След това модулът за генерация AUTO (евентуално) сравнява откритията на системата с тези на другия агент и генерира цялостна научна статия, съдържаща заглавие, уводна част, открити от програмата универсали, (евентуално) сравнение с тези на другия агент и заключение. Важни открития на програмата представляват направените в класическата словоредна база от данни на Дж. Гринберг, един от

водещите лингвисти на 20-ти век. Системата е открила както много повече универсали, така и някои грешки в предложените от Гринберг. Системата е породила няколко такива текста в областта на словореда, два от които са публикувани без човешка редакция и без предварителното знание на редакцията за машинния произход на статиите, в сп. „Съпоставително езикознание" (вж. [11, 14] от списъка на всички публикации, прил. **Публикации общо.pdf**). Списък от 50 фонологични универсали са включени във вида, породен от системата, в авторитетния Universals Archive, University of Konstanz, а повече от 100 в Глава 5 на монографията ми [М-2].

В [8] системата UNIVAUTO е използвана за намиране на „фонологични особености" посредством генерация на контрапримери на статистически универсали, търпящи много малък на брой изключения. Това представлява една интересна, но неизвестна преди това интерпретация на понятието „статистически универсал".

Статия [10] се занимава с проблема за простотата на описание на една лингвистична типология като множество от универсали. Показва се, че всяка такава типология може да се опише с множество от нестатистически универсали (в противовес на приетото в лингвистиката) и се описва компютърен метод (MINTYP) за намиране на най-малкото описващо множество, комбиниращ модула UNIV с модул за намиране на минимално покритие. Системата MINTYP е намерила минимални описания значително различаващи се от тези, предложени от авторитетни лингвисти като Гринберг и Хокинс. Статията е публикувана във водещото списание *Linguistic Typology* (официален орган на Асоциацията по лингвистична типология).

Друго приложение на системата UNIVAUTO [4] е върху фонологични данни, посветено на изследване на характерния за лингвистиката тип импликации от вида „Ако фонема А, то фонема В". Изследвайки същите данни като известния фонолог Мадиесън (Университета на Калифорния в Лос Анджелис), системата – както и при предишните ни подобни изследвания – е намерила както значително по-голям брой универсали, така и проблеми в предложените от Мадиесън (напр. отсъствие на статистическа значимост). Получените повече от 140 универсала са машинно изследвани, за да се стигне до интересен лингвистичен принцип, отнасящ се до структурата на фонемата-антецедент и фонемата-консеквент в една импликационна универсала. Статията е публикувана в авторитетното списание *Folia Linguistica* (официален орган на Европейската асоциация по лингвистика).

Статия [1] показва логически проблеми при формулировката на универсали в примери на известни лингвисти от литературата и е публикувана в същото списание.

[20] прави оценка на UNIVAUTO като първата програма, която генерира научна статия за своите открития. Предложени са критерии, които да характеризират знанията намирани от една успешна програма за откритие (новост, интересност, достоверност, разбираемост, преносимост и евристичност), и е показано, че UNIVAUTO отговаря добре на тези критерии.

Статия [12] е обзорна и прави преглед на четири програми за откритие на автора, като се хвърля поглед и върху бъдещото развитие на областта на машинно откритие в лингвистиката.

МОНОГРАФИЯ [М-2] има обем от 330 стр. и обхваща 8 глави, предговор и 2 индекса (предметен и на използваните езици) и е публикувана от водещото международно издателство Equinox, базирано в Лондон. Тя има за цел да обобщи многогодишната ми работа върху машинното откритие в лингвистиката и да послужи като въведение в тази нова област в лингвистиката. Макар и през годините да има немалко работи, използващи машина за решаване на отделни лингвистични проблеми, тези работи се появяват в различни (често много специализирани и недостъпни)

издания, те не дават цялостна картина на областта и което е най-важно, най-често се ограничават до отделни стандартни проблеми (намиране на авторство, историческа граматика или диалектология) или пък тези проблеми нямат творчески, а само рутинен характер. В Глава 1 на книгата сравнително подробно (вкл. в исторически план) се разглежда какво е откритие в лингвистиката и факторите, подпомагащи откритието (интуиция, случайност, решаването на проблеми). Решаването на проблеми се показва като важен компонент, който може (поне частично) да се автоматизира, като оригиналната задача се сведе до една или повече от повтарящи се в различните области задачи като намиране на прилики/разлики, образуване/ревизиране на понятия, разграничаване на множество от понятия, формиране/ревизиране на таксономии, намиране на индуктивни закони и др. Освен тези генерични научни задачи, лингвистът обикновено решава и мета-научни задачи като намиране на всички или най-просто решение, проверка на неговата достоверност и пр. Глави 2-7 представляват илюстрации на изградената в първата глава концептуална схема на базата на някои от предишните ни изследвания [4, 5, 7, 8, 10, 13, 14], които са допълнени и обобщени в новата концептуална рамка. Глави 2 и 3 са посветени на генеричната научна задача за разграничаване на множество от понятия, която се илюстрира с разнообразни задачи от лексикологията (компонентен анализ на термините за родство), фонологията (анализ по дистинктивни признаци; профилиране на отделни езици), езиковата патология и др., като описаните системи решават и мета-научните задачи за осигуряване на всички или на най-простите решения. Глави 4 и 5 разглеждат генеричната научна задача за откриване на индуктивни закони, която се илюстрира с примери от словореда, фонологията, родствените модели и др., като системите решават и мета-научните задачи за осигуряване на достоверност, простота и пълнота на решенията. Глава 6 третира откриването на най-просто описание на типология и обединява научната задача за индуктивна генерализация с мета-научните задачи за простота и пълнота. Глава 7 разглежда задачата за намиране на статистически значими прилики и я илюстрира с откритието на лингвистична връзка между Южна Америка и Океания, като също подробно се спира на лингвистичните аргументи, които подкрепят хипотезата. В заключителната Глава 8 резултатите се обобщават и се предлага схема на изследователския цикъл в лингвистиката.

Като цяло представената монография дава нов, компютърен поглед върху лингвистичната методология. Вместо лингвистичните методи да се свързват с отделните лингвистични дисциплини (фонология, семантика, типология, историческа граматика и т.н.), както се прави до сега, те се разглеждат като генерични научни и мета-научни задачи, до които се свеждат конкретните проблеми, които се появяват в тези дисциплини. Това е първата книга по машинно откритие в лингвистиката и тя е положително рецензирана в сп. “Computational Linguistics” (официален орган на Международната асоциация по компютърна линвистика). Отбелязана е в Cyberling Wiki и Linguist List.

С. Компютъризирано изучаване на еволюцията на езика (7): М-1, 2, 3, 5-7, 19

Еволюцията на езика, изучаването на неговата предистория/история и разпространение и свързаните с това миграции, представлява важна тематика в последните десетина години, когато се появяват статии по въпроса в списания като *Nature* и *Science*. За разработването на такава тематика беше създаден „Института по еволюционна антропология - Макс Планк”, където авторът беше поканен да подпомогне с

компютърни средства съвместна работа на лингвисти и генетици. Представените по-долу работи са свързани с подобни въпроси.

С1. Хипотезата за езикова връзка между кайнганг (Бразилия) и океанските езици
Идеята за потенциална езикова връзка между езиците от семейство кайнганг (хокленг и кайнганг, говорещи се в Бразилия) и океанското семейство (голям брой езици, говорещи се в Тихия океан) е компютърно породена посредством програмата на автора RECLASS, която намира в представителни езикови бази от данни на статистически значими прилики между езици, които принадлежат – според съвременните класификации – към *различни* езикови семейства. Подобна хипотеза не е била изказвана в лингвистичната литература преди.

В [6] е описана програма, която сравнява по двойки стандартни списъци от 100 думи (списъци на Суодеш). Програмата е намерила, че, сравнявайки бразилския език хокленг с океанските малайски, тагалог, фиджийски, самоански и хавайски, тези сравнения показват статистически значими лексически прилики.

[5] разглежда програмата RECLASS, довела до предложената хипотеза. В статията са показани значимите прилики на „родствени модели" (kin term patterns) между езици от двете семейства. По-подробно са отбелязани фонологични и граматични сходства, а за около 40 сравнявани лексеми от хокленг и кайнганг от една страна, и хавайски и маори от друга, са формулирани фонетични съответствия между езиците, стандартният лингвистичен начин за доказване на генетична връзка между езици.

В [19], доклад на 19-ия Световен конгрес на лингвистите, приет от тогавашния редактор на сп. "Language" (официален орган на Американската асоциация по лингвистика), са приведени по-нататъшни поддържащи хипотезата данни от фонологията, граматиката и лексиката и горните резултати са обобщени.

Статия [3] разглежда подробно термините за родство (22 термина) в сравняваните семейства от езици и показва звуковите съответствия между тях, което представлява силна подкрепа на предложената хипотеза. Изложени са и други подкрепящи аргументи от публикувани работи по генетика и антропология.

Предложената от автора хипотеза има важни следствия не само за лингвистиката, но и за дисциплини като генетика, антропология, археология и човешка предистория и миграции. Надеждността на хипотезата е оценена положително в Wikipedia, в рецензия на книгата ми [М-2] в списанието "Computational Linguistics" и от водещия специалист по разглежданите бразилски езици Урсула Виземан (вж. по-долу във **- впечатляващи коментари на учени с висок международен престиж**).

С2. Други
Следващите две статии най-общо имат отношение към хипотезата за съществуване на положителна корелация между големината на фонемния инвентар на езиците и броя на хората, говорещи тези езици. И двете са написани по покана на редактора на водещото списание *Linguistic Typology*.

[7] е коментар на статия в същото списание на известния социолингвист Питър Тръджил, който предлага (усложнен) вариант на горната хипотеза. В [7] се привеждат резултати от изследване на големините на фонемния инвентар и броя на хората, говорещи над 400 езика, от които тезата ясно се опровергава. Работата е коментирана в редица блогове и статии, вкл. и в *Proceedings of the US National Academy of Sciences, PNAS.*

[2] е коментар на статия в *Science* на генетика Куентин Аткинсън, който се опитва да докаже произход на човешкия език от Африка, интерпретирайки по-надежден вариант на горната хипотеза като т. нар. “serial founder effect”. В работата се показват някои слаби елементи в аргументацията на Аткинсън, показват се основните звена, които подобна аргументация би следвало да съдържа, както и различни други (разумни) интерпретации на данните, което поставя под сериозно съмнение верността на хипотезата.

МОНОГРАФИЯ [М-1] има обем от 177 стр., обхваща 6 глави и 2 приложения, и е публикувана от известното специализирано в лингвистика издателство LINCOM EUROPA, базирано в Мюнхен. Книгата продължава изследванията ни по автоматизирано откритие в лингвистиката в областта на историческата и сравнителна лингвистика и има за цел да профилира езиковите семейства (стандартна задача в дисциплината), което тя прави в термините на техните родствени модели (kin term patterns). Глави 1 и 2 имат уводен характер, разглеждат главните проблеми в областта на изучаването на родствените модели от лингвистична и антропологична гледна точка и различни подходи към тяхното решаване, както и въвеждат основните понятия. Особено внимание се отделя на базата от данни на родствени модели, създадена от известния антрополог Дж. Мърдок, обхващаща 566 езика от целия свят. Глава 3 въвежда системата MPD (Maximally Parsimonious Discrimination), с помощта на която се извършва профилирането, което може да използва както елементарни атрибути (родствени модели), така и производни атрибути (комбинация от родствени модели), в случая когато елементарните са недостатъчни за разграничение. Глави 4 и 5 представляват основният принос на книгата. В Глава 4 езиците от базата от данни на Дж. Мърдок се групират в езикови семейства според *Ethnologue* (общоприетата класификация) и нашата програма MPD (Maximally Parsimonious Discrimination) се използва те да се профилират, като всяко семейство се разграничава от всички останали по най-икономичен начин с елементарни и (евентулно) производни атрибути. В главата се дават профилите на всички 64 езикови семейства. В Глава 5 се разглеждат различни начини, по които откритите профили да се използват в изследвания по историческа и сравнителна лингвистика и типология. Основен извод е, че профилите на родствените модели могат да се считат като важни индикатори на генетична принадлежност. Най-важни в това отношение се оказват моделите за братя/сестри, техните съпруги/съпрузи и деца. В книгата се разглеждат критично значителен брой хипотези относно генетична принадлежност, предложени в литературата, и те са оценени в светлината на откритите профили.

Книгата е цитирана в сп. “Science” и реферирана в InterSciWiki и Linguist List.

По чл.2 т.6 от Правилника на ИМИ за приложение на ЗРАСРБ:

- впечатляващи коментари на учени с висок международен престиж (през последните 3 години):

E.J.Smith. Review of *Machine-Aided Linguistic Discovery: An Introduction and Some Examples* by Vladimir Pericliev. *Computational Linguistics* 2010, 36(4): 784-787, относно Публикация [М-2]:

> “The subtitle of Vladimir Pericliev’s book, *An Introduction and Some Examples*, is a succinct and accurate description of its contents. <..> Pericliev’s essential point is a valid one: Machine-aided discovery has a tremendous untapped potential for analyzing data sets which are too large to be amenable to human

inspection. The success of this approach is best exemplified by his machine-aided discovery of a possible genetic relationship which would otherwise have eluded human discovery."

James Winters. The great mystery of the vanishing phoneme. in *A replicated Typo: Language, its evolution and anything in-between.* January 26, 2012 (http://replicatedtypo.com/the-great-mystery-of-the-vanishing-phonemes/4525.html) относно Публикация [2]:

"For now, I'll just provide a brief list of the authors (and their papers) that critique Atkinson's original article, which, if you know your linguists, pretty much reads like an A-list of the big names in the field <..>:

Joan Bybee: How plausible is the hypothesis that population size and dispersal are related to phoneme inventory size? Introducing and commenting on a debate.

Peter Trudgill: Social Structure and phoneme inventories.

Mark Donohue and Johanna Nichols: Does phoneme inventory size correlate with population size?

Östen Dahl: Are small languages more or less complex than big ones?

Søren Wichmann, Taraka Rama, Eric W. Holman: Phonological diversity, word length, and population sizes across languages: The ASJP evidence.

Richard Sproat: Phonemic diversity and the Out-of-Africa theory.

Claire Bowern: Out of Africa? The logic of Phoneme Inventories and Founder Effects.

Vladimir Pericliev: On phonemic diversity and the origin of language in Africa.

Don Ringe: A pilot study for an investigation into Atkinson's hypothesis. икаци

Keren Rice: Athabaskan languages and serial founder effects.

Bill Ross and Mark Donohue: The many origins of diversity and complexity in phonology.

Ian Maddieson, Tanmoy Bhattacharya, Eric D. Smith and William Croft: Geographical distribution of phonological complexity.

Florian T. Jaeger, Peter Graff, William Croft and Daniel Pontillo: Mixed effect models for genetic and areal dependencies in linguistic typology.

Quentin D. Atkinson: Linking spatial patterns of language variation to ancient demography and population migrations."

Статия за езика кайнганг в Wikipedia at http://en.wikipedia.org/wiki/Kaingang_language относно Публикации [3, 5, 6]:

"In the 1960s, because of a missionary interest (conducted by the Summer Institute of Linguistics (SIL)), the language was studied by Ursula Wiesemann. Wiesemann proposed an orthography for the language, which is still in use (in spite of some troubles). <..>. Further investigation of the language by Vladimir Pericliev has suggested a convincing lexical affinity with Polynesian languages within the Austronesian linguistic family. Several theories have been proposed for this apparent similarity - from cultural diffusion to common mother-tongue - although further study is required."

Dr. Ursula Wiesemann, водещият специалист по кайнганг и цитирана по-горе, в лична кореспонденция (вж. прил. **Други докум2.pdf**) относно Публикация [19]:

"Thank you for the excellent presentation of these most surprising facts. I dont think that anyone can show that you would be wrong <..>

I am still amazed at all of this and am wondering about the implications myself. <..>

As for me, I still marvel at your findings, and don't really know what conclusions to draw. And I congratulate you for having found this."

Prof. Emeritus Bent Sigurd, Lund University, именит (компютърен) лингвист (вж. http://sv.wikipedia.org/wiki/Bengt_Sigurd) в лична кореспонденция (вж. прил. **Други докум2.pdf**) относно Публикация [М-2] и хипотезата за лингвистична връзка между Южна Америка и Океания:

" Your book has arrived and I am very impressed. <..> I will read the book and try to learn.

<..> I have now read part of chapter 7 and noted your even more convincing argumentation. The so called specialists cannot ignore the argument and the interesting problem much longer. I even think the media will note and report your discoveries. <..>
I suggest the following headline: Heyerdahl was right! But the direction was wrong! says Bulgarian mathematician...."

Prof. Emeritus Ian Maddieson, University of California at Berkeley, водещ фонетик и типолог (вж. http://en.wikipedia.org/wiki/Ian_Maddieson) в лична кореспонденция относно покана за рецензия (вж. прил. **Други докум2.pdf**):

"I'm impressed, as always, by the range of your work."

Prof. Douglas White, University of California at Urvine, водещ математически антрополог (вж. http://en.wikipedia.org/wiki/Douglas_R._White) относно Публикация [М-1] (вж. прил. **Други докум2.pdf**):

"I have examined your book with great interest.
I referenced it on my wiki site by opening a page for you at http://intersci.ss.uci.edu/wiki/index.php/Vladimir_Pericliev"

Jacob Fitisemanu (University of Utah Medical School) в лична кореспонденция (вж. прил. **Други докум2.pdf**) относно Публикация [6] и др.:

"I recently read your paper "Significant Lexical Similarities between a Language of Brazil and Some Languages of Southeast Asia and Oceania," and was very intrigued by your research.
<..> I am very grateful for your treatment of this unique topic and I look forward to reading more of your research.
<..> Thank you again for pointing out your other papers and for conducting the work that you are doing, it is very much appreciated!"

Silo Chin (Managing editor of Journal of Universal Language) в лична кореспонденция относно покана за написване на статия (вж. прил. **Други докум2.pdf**):

"Due to the fact that the Journal of Universal Language is now poised for take-off, your contribution to the journal would be momentous and certainly would set the tone for a successful future for the journal."

Prof. Victor de Munck (Head of Anthropology Dpt., SUNY New York) в лична кореспонденция (вж. прил. **Други докум2.pdf**) относно желанието му за съвместна работа в София по темата за родствени термини (вж. Публикации [М-1, 14] от представените и [М-3] от общия списък) и кандидатстването му за стипендия от Center for Advanced Study, Sofia:

"I applied for a number of reasons: my urge to go places; extend my knowledge of eastern europe; and, perhaps most importantly it might give me an opportunity to work with you.
<..>It's what i've come to expect from you kindness and intelligence. i'm very interested in computational modeling so i'm hoping that some of your knowledge in this field will rub off on me."

- трудове на други учени, публикувани в авторитетни издания, които съществено използват резултати на кандидата:

Компютърните ни програми за икономично, разбираемо и приближено профилиране на множество от класове, описани в съвместните Публикации [9, 11, 14, 23] от списъка на представените, са използвани за обработка на бази от данни в различни области (биология, психология и химия) в следните авторитетни издания:

1. Radwan E. Abdel-Aal, Mona R. E. Abdel-Halim, and Safa Abdel-Aal. 2006. Improving the classification of multiple disorders with problem decomposition. ***Journal of Biomedical Informatics*** 39: 612–625. **[IF 1.719]**
2. MacWhinney, B. B., Feldman, H., Sacco, K., and Valdés-Pérez, R. 2000. Online measures of basic language skills in children with early focal brain lesions. ***Brain and Language*** 71: 400–431. **[IF 3.162]**
3. Zeigarnik, A. V., R. E. Valdés-Pérez, and J. Pesenti. 2000. Comparative properties of transition metal catalysts inferred from activation energies of elementary steps of catalytic reactions. ***Journal of Physical Chemistry*** 104: 997–1008. **[IF 4.524]**
4. Valdés-Pérez, R. E., A. Zeigarnik, and J. Pesenti. 2002. Science: similarities and differences among catalysts: Clustering and profiling diverse data on chemical reactions. In *Handbook of Data Mining and Knowledge Discovery*, pp. 967–972. **Oxford University Press**, Inc. New York, NY.

Съвместната ни работа с Раул Валдес-Перез (вж. http://www.linkedin. com/ in/ valdesperez), отразена в Публикации [9, 11, 14, 23] от списъка на представените и Публикации [41, 43, 47] от общия списък по създаване на програми за ефективно профилиране на множество от класове е продължена от него с патентоване на програма за ефективно профилиране на индивиди (използвана частично в [9] от представените публикации) и съосноваването през 2000 г. на много успешната компания Vivisimo (вж. http://en.wikipedia.org/wiki/Vivisimo), която е закупена през май 2012 от IBM.

Съвместните Публикации [9, 11, 14, 23] от списъка на представените са съществено използвани в концептуален план за дефиниране на понятието „успешна система за научно откритие", в статия, публикувана в най-авторитетното списание по изкуствен интелект, а [11] в идеята за профилиране на алгоритми с цел техния избор, съответно в:

1. R. E. Valdés-Pérez. 1999. Principles of human-computer collaboration for knowledge discovery in science. ***Artificial Intelligence*** 107: 335- 346 **[IF 2.511]**.
2. N. Ramakrishnan and R. E. Valdés-Pérez. 2000. Note on Generalization in Experimental Algorithmics. ***ACM Transactions on Mathematical Software*** 26(4): 568-580. **[IF 1.9]**

Данните, описани в Публикация [7] от списъка на представените, са използвани за по-нататъшна обработка в статия в официалния орган на Американската асоциация по лингвистика:

1. Jennifer Hay and Bauer. 2007. Phoneme inventory size and population size. ***Language*** 83: 388–400. **[IF 2.026]**

Около 50 фонологични универсали, открити от системата UNIVAUTO, описана в Публикации [4, 7, 8, 20, 21] от списъка на представените, се съдържат в авторитетния архив на универсали, поддържан от редактора на водещото списание *Linguistic Typology*:

1. Franz Plank et.al. The UNIVERSALS ARCHIVE, http://typo.uni-konstanz.de/archive/intro/

Различни аспекти на представените публикации са разгледани повече или по-малко подробно в различни авторитетни източници. По-важни разглеждания, изразени в цитирания, са:

В списания с много висок импакт фактор в науката въобще (вж. прил. **Цитати.pdf**)

1. Charles Kemp and Terry Regier. 2012. Kinship categories across languages reflect general communicative principles. *Science* 25 May 2012: Vol. 336 no. 6084 pp. 1049–1054. (**IF 31.36,** цитира две работи, [М-1] и [16] от представени публ.)

2. Daniel Nettle. 2007. Language and genes: A new perspective on the origins of human cultural diversity. *Proceedings of the US National Academy of Sciences, PNAS*, June 26, 2007, vol. 104, no. 26: 10755–10756. (**IF 9.771**, цитира [7] от представени публ.)
3. C. Perreault and S. Mathew. 2012. Dating the Origin of Language Using Phonemic Diversity. *PLoS ONE* 7(4): e35289. doi:10.1371. (**IF 4.441**, цитира [2] от представени публ.)

В книги в авторитетни издателства (избрани от общо 29) (вж. прил. **Цитати.pdf**)

1. Steven Moran. 2012. Using Linked Data to create a typological knowledge base. In Chiarcos et al., Linked Data in Linguistics, pp. 129–138, **Springer**.
2. Lyle Campbell and William J. Poser. 2008. Language Classification: History and Method. **Cambridge University Press**.
3. Peter Trudgill. 2011. Sociolinguistic Typology: Social Determinants of Linguistic Complexity. Oxford: **Oxford University Press**.
4. Eric Raimy and Charles Cairns, 2009. Contemporary Views on Architecture and Representations in Phonological Theory. MIT, Cambridge, MA: **MIT Press**.
5. Gabriel Altmann, Reinhard Köhler, R. Piotrowski (eds.). 2005. Quantitative methods in typology. Quantitative Linguistics: An International Handbook. (HSK). Berlin: **Mouton de Gruyter**.
6. Laxmi Parida.2007. Pattern Discovery in Bioinformatics: Theory and Algorithms. **Chapman & Hall**, Mathematical & Computational Biology Series.
7. Segev, Aviv. 2006. Identifying the Multiple Contexts of a Situation. Lecture Notes in Computer Science 3946: 118–133. **Springer**.
8. H. Russel Bernard. 2006. Research Methods in Anthropology. **Altamira Press**, 4th edition.
9. Madeline McClenney-Sadler. 2007. Recovering the Daughter's Nakedness: A Formal Analysis of Israelite Kinship Terminology and the Internal Logic of Leviticus 18. London: **T&T Clark**.
10. Lorenzo Magnani. 2009. Abductive Cognition. **Springer**.
11. Langley P. 1998. The computer-aided discovery of scientific knowledge. In Carbonell JG, Siekmann J, editors. Lecture Notes in Artifical Intelligence. Vol. 1532, pp. 25–39. New York: **Springer**.
12. Langley P. 2001. The computational support of scientific discovery. In Carbonell JG, Siekmann J, editors. Lecture Notes in Artifical Intelligence. Vol. 2049. New York: **Springer**; 2001. pp. 230–48.
13. Philip Good, 2005. Resampling Methods: 3rd edition: Practical Guide to Data Analysis, **Birkhauser: Boston**.
14. Martin Haspelmath and Ekkehard König and Wulf Oesterreicher and Wolfgang Raible] (eds.). 2001. Language Typology and Language Universals: An International Handbook. (Handbücher zur Sprach- und Kommunikationswissenschaft) Vol. 1–2. Berlin: **de Gruyter**, 1856 pp.
15. Magnani, L. 2001. Abduction, Reason and Science. **Kluwer Academic/Plenum Publishers**.
16. G. Palioras. 2001. Machine Learning and Its Applications: Advanced Lectures. **Springer**.
17. S. Nirenburg and V. Raskin, Ontological Semantics. **MIT Press**, 2004.
18. Maxwell, Dan, Klaus Schubert, and Toon Witkam (eds). 1988. New Directions in Machine Translation. **Dordrecht, Holland: Foris**.
19. S. Nirenburg, J. Carbonell and M. Tomita. 1994. Machine Translation: A Knowledge-Based Approach. **Morgan Kaufmann Publishers** Inc. San Francisco.
20. S. Nirenburg, H. Somers, Y. Wilks. 2003. Readings in Machine Translation. **The MIT Press**.
21. S. Nirenburg. 1992. Machine Translation: A Knowledge-based Approach. **Morgan Kaufmann Publishers**.
22. G. Hirst. 1992. Semantic Interpretation and the Resolution of Ambiguity. **Cambridge University Pres**s.
23. Melvin Konner. 2010. The Evolution of Childhood. **Harvard University Press**.
24. R. Valdés-Pérez. 2004. Recollections from 15 years of monthly meetings. In Models of a Man: Essays in Memory of Herbert Simon. **The MIT Press**.

В дисертации в авторитетни университети (избрани от общо 17) (вж. прил. **Цитати.pdf**)

1. Sinnemäki, Kaius. 2011. Language Universals and Linguistic Complexity: Three case studies in core argument marking, **University of Helsinki**, 2011, Doctoral Dissertation.
2. Shou-de Lin. 2006. Modeling, Searching and Explaining Abnormal Instances in Multirelational Networks. PhD Dissertation, **University of Southern California**.
3. Deept Kumar. 2007. Redescription Mining: Algorithms and Applications in Bioinformatics. PhD Dissertation, Department of Computer Science, the Faculty of the **Virginia Polytechnic Institute and State University**.
4. Jonathan Schug. 2005. Integrating Gene Expression Signals with Bounded Collection Grammars. PhD Dissertation, Computer Science, **University of Pennsylvania**.
5. Sheldon Chow. 2011. Heuristics, Concepts, and Cognitive Architecture : How the Mind Works. PhD Dissertation at the **University of Western Ontario** London, Ontario, Canada, 2011.
6. Marco Kuhlmann. 2007. Dependency Structures and Lexicalized Grammars. PhD Dissertation, Naturwissenschalich-Technischen Fakultäten der **Universität des Saarlandes**.
7. Hadar Shemtov. 1997. Ambiguity Management in Natural Language Generation. Doctoral Dissertation, **Stanford University**, Stanford.
8. M. Malik. 2010. Méthodes et outils pour les problèmes faibles de traduction. PhD Dissertation, **University of Grenoble**.

По чл.3 от Правилника на ИМИ за приложение на ЗРАСРБ:

- ръководство и участие в международни и национални научноизследователски проекти:

1. Индивидуален проект *Computerized linguistic methods in support of evolutionary genetics investigations* (EC Marie Curie Fellowship MCFI-2001-00689) с „Института по еволюционна антропология - Макс Планк" (гост-изследовател в департаментите по генетика и лингвистика за 7 месеца; приемащ-учен проф. Марк Стоункинг). (2002)
2. Индивидуален проект *Generic tasks of scientific discovery* (Supplement to NSF grant #IRI-9421656 from the USA NSF Division of International Programs) с Университета „Карнеги Мелън" (гост-изследовател в департамента по информатика за 6 месеца; приемащ учен д-р Раул Валдес-Перез). (1997-1998)
3. Проект *Компютърни средства в помощ на изследователската работа на лингвиста* (дог. МИ-1511/2005 с НФНИ) (ръководител на договор). (2002-2007)
4. Проект *Аспекти на машинното откритие и приложение в лингвистиката* (дог. И-813-95 с НФНИ) (ръководител на договор). (1997-2000)
5. Индивидуален проект *Машинно извличане на знания в лингвистиката* с ИМИ-БАН. (1997-)

- участие в програмни комитети на научни мероприятия:

1. Sixth International Conference on Discovery Science, Sapporo, 2003.
2. Fifth Conference of the European Chapter of the Association for Computational Linguistics, Berlin, 1991.

3. Fourth Conference of the European Chapter of the Association for Computational Linguistics, Manchester, 1989.

- участие с доклади в международни и национални научни форуми (избрани)

1. Pericliev, V. 2002. A linguistic discovery system that verbalizes its discoveries. *COLING02, Proceedings of the 19th International Conference on Computational Linguistics*, Taipei, Taiwan, August 24-September 1, pp. 1258-62.
2. **Pericliev, V**. and R. Valdes-Perez. 1998. A procedure for multi-class discrimination and some linguistic applications. *COLING98, Annual Meeting of the Association of Computational Linguistics & Proceedings of the 17th International Conference on Computational Linguistics*, Montreal, Quebec, Canada, August 10-14, pp. 1036-42.
3. Pericliev, V. 1998. The prospects for machine discovery in linguistics. ICDC, 1st International Conference on Discovery and Creativity. 14-16 May, 1998, Gent, Belgium.
4. Pericliev, V. 1997. From the principle of projectivity and partial orderings to sophisticated linearization rules. Invited paper for the "Constraining dependency grammar" panel, *16th International Congress of Linguists*, Paris 1997.
5. **Pericliev, V**. and R. Valdes-Perez. 1997. A discovery system for componential analysis of kinship terminologies. *16th International Congress of Linguists*, Paris 1997.
6. Pericliev, V. 1996. Learning linear precedence rules. *COLING96, 16th International Conference on Computational Linguistics*, pp. 883-888, 5-9 August, Copenhagen, 1966.
7. Pericliev, V. 1995. Empirical discovery in linguistics. In *Systematic Methods of Scientific Discovery. Spring Symposium of the American Association for Artificial Intelligence*. Stanford University, California, March 1995, pp. 68-73.
8. **Pericliev, V**., S. Brajnov, I. Nenova. 1988. Hinting by paraphrasing in an instruction system. *COLING88, Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, August 1988, pp. 507-511.
9. Pericliev, V. 1987. Are all sentences with constructional homonymity ambiguous?, *Proceedings of the 14th International Congress of Linguists*, Berlin, August 1987, 1032-1034.
10. Pericliev, V. 1987. Heuristics in a linguistic investigation, *Abstracts of the 8th International Congress of Logic, Methodology and Philosophy of Science*, Moscow, August 1987, 502-504.
11. **Pericliev, V**. and I. Ilarionov. 1986. Testing the projectivity hypothesis, *COLING86, Proceedings of the 11th International Conference on Computational Linguistics*, Bonn, August 1986, pp. 56-58.
12. Pericliev, V. 1984. Handling syntactical ambiguity in Machine Translation. *COLING84, Proceedings of the 10th International Conference on Computational Linguistics*, Stanford, California, August 1984, pp. 521-524.

- членство в авторитетни творчески и/или професионални организации в съответната научна област:

1. ACL (Association for Computational Linguistics): 1983 – 2000
2. SLE (Societas Linguistica Europaea): 2008 -

- участия в редколегии на научни издания:

GLOSSA: An Ambilingual Interdisciplinary Journal: 2010-

- авторитетни отзиви:

Монографията [М-2] от списъка на представените е рецензирана положително в сп. "Computational Linguistics", най-авторитетното издание в областта с един от най-високите импакт фактори (**IF 2.971**) в лингвистиката въобще (вж. E. J. Smith. Review of *Machine-Aided Linguistic Discovery: An Introduction and Some Examples* by Vladimir Pericliev. *Computational Linguistics* 2010, 36(4): 784-787). Вж. също точка „Впечатляващи коментари на учени с висок международен престиж".

- създаване на ново направление в науката:

Монографията [М-2] от списъка на представените е първата книга по машинно откритие в лингвистиката.

- изнасяне на лекции в чуждестранни университети:

1. Max Planck Institute for Evolutionary Anthropology (с Марк Стоункинг (генетика) и Бил Крофт (лингвистика)) (2002)
2. Carnegie Mellon University (с Раул Валдес-Перез (информатика)) (1998)

- експертна дейност в международни организации:

Външен рецензент на хабилитацията на д-р Раул Вардес-Перез в департамента по информатика на Университета „Карнеги Мелън".

- дейности, свързани с научното развитие на ученици:

Подготовка и провеждане на олимпиади по математическа лингвистика през периода 1986-1989.

- публикации и други дейности по популяризирането на науката:

1. Периклиев, Вл. Конкурсна задача по математическа лингвистика. сп. „Математика", 1986. 5.
2. Периклиев, Вл. Конкурсна задача по математическа лингвистика. сп. „Математика", 1986. 6.
3. Периклиев, Вл. Конкурсна задача по математическа лингвистика. сп. „Математика", 1987. 1.
4. Периклиев, Вл. Конкурсна задача по математическа лингвистика. сп. „Математика", 1987. 7.
5. **Периклиев, Вл**., И. Периклиева. Увод в организацията на експертните системи. В кн. „Експертни системи". Наука и изкуство. София, 1989. 163-202 (превод от английски).
6. **Периклиев, Вл**., И. Периклиева. Методи за пораждане на обяснения. В кн. „Експертни системи". Наука и изкуство. София, 1989. 239-265 (превод от английски).
7. **Периклиев, Вл**., И. Периклиева. Търговски програмни средства. В кн. „Експертни системи". Наука и изкуство. София, 1989. 359-405 (превод от английски).
8. **Периклиев, Вл**., И. Периклиева. Експертни системи: действащи системи и изследователската литература. В кн. „Експертни системи". Наука и изкуство. София, 1989. 407-442 (превод от английски).
9. Периклиев, Вл. (Редактор) N. Bozhinov. Convolutional Representations of Commutants and Multipliers, Sofia, 1988. House of the Bulgarian Academy of Sciences.

юни 2012 г. Подпис: