

**Николай СТАНКОВ**

Великотърновски университет „Св. св. Кирил и Методий“, България

## **МЕТОДИКА НА КВАНТИТАТИВНИТЕ МОРФОЛОГИЧНО-ТИПОЛОГИЧНИ ИЗСЛЕДВАНИЯ. ПРОБЛЕМИ И РЕШЕНИЯ**

Nikolay STANKOV

St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria

### **THE METHODOLOGY OF QUANTITATIVE MORPHOLOGICAL-TYPOLOGICAL STUDIES. PROBLEMS AND SOLUTIONS**

**Резюме:** Много от съвременните изследвания в областта на квантитативната морфологична езикова типология, които са базирани върху работата на Джоузеф Грийнберг, се сблъскват с няколко важни методологически проблема: липсата на универсална и изчерпателна дефиниция за понятието „дума“; твърде общата употреба на понятията „аналитичен“ и „синтетичен“; недоизяснения статус на някои класове думи (напр. предлози и съюзи) като свободни граматични маркери, имащи отражение върху индекса на аналитичност; потенциала за изкривени резултати при анализа и сравнението на аналитичността и синтетичността на аглутинативни и флективни езици. Настоящата разработка дава предложения за това как тези проблеми биха могли да бъдат решени или поне смекчени, за да могат квантитативните типологически изследвания на морфологично ниво да бъдат по-релевантни и информативни на фона на високите изисквания за прецизност и задълбоченост в съвременната наука.

**Abstract:** Many of the current studies in the field of quantitative morphological linguistic typology, which are based on the work of Joseph Greenberg, face several important methodological problems: the lack of a universal and comprehensive definition of the concept 'word'; the overly general use of the terms 'analytic' and 'synthetic'; the underclarified status of some word classes (e.g prepositions and conjunctions) as free grammatical markers having an impact on the index of analyticity; the potential for skewed results in the analysis and comparison of the analyticity and syntheticity levels of agglutinative and inflectional languages. This paper provides suggestions for solving or at least alleviating these problems, so that quantitative typological research at the morphological level may be more relevant and informative in the face of the high demands for precision and thoroughness in modern science.

**Ключови думи:** *квантитативни типологически изследвания, аналитизъм, синтетизъм, граматичност, граматични маркери*

**Keywords:** *quantitative typological research, analyticity, syntheticity, grammaticity, grammatical markers*

## 1. Въведение

Квантитативната морфологична типология е клон на лингвистиката, който изследва, анализира и сравнява различните типове езици на морфологично равнище с помощта на количествени методи, заимствани от сферата на математиката и статистиката. Този вид изследвания могат да имат значителен принос в областта на контрастивната и диахронната лингвистика, тъй като дават емпирични доказателства за това, как и до каква степен езиците се различават един от друг по отношение на своя морфологичен тип, както и в каква посока се развиват исторически. Морфологичната езикова типология е дисциплина, която води началото си от 19. век и трудовете на братята Фридрих и Аугуст Шлегел (Schlegel, F. 1808, Schlegel, A. 1818). От този период произлиза и разграничението на езиците на изолиращи, аглутинативни и флективни – понятия, които междувременно са се наложили като стандартни в езикознанието. Други две типологични категории, които влизат в употреба през същия период, са тези на аналитичните и синтетичните езици. Първоначално те са подкатегории на флективните езици, но през 20. век понятията започват да се употребяват и във връзка с останалите видове езици.

Влиятелните в лингвистиката трудове на Едуард Сапир (Sapir 1921) и Джоузеф Грийнберг (Greenberg 1960) допринасят за популярността на тези термини през 20. век, като те остават актуални и в съвременето. В курсовете по лингвистика обикновено те се въвеждат още в първия семестър от обучението на студентите. Често се използват като дескриптивни понятия с цел обща типологична класификация на езици, междуезикови сравнения и определяне посоката на езиковия развой от морфологично-типологична гледна точка. Твърди се например, че „по отношение на граматиката английският език се характеризира с пределна степен на аналитичност“ (Братанова 2021: 127) или че българският език „от преобладаващо синтетичен език, който притежава граматическата категория падеж и сложна склонитбена система от седем падежа в единствено, двойствено и множествено число при имената, днес той е безпадежен и като цяло е преобладаващо аналитичен“ (Петров 2021: 150). Определения като „пределно аналитичен“ или „преобладаващо аналитичен“ обаче са критикувани още през епохата, следваща един от първите опити за подобна типологична класификация на световните езици, предложена от Едуард Сапир в придобилата широка популярност за времето си книга *Language* (Sapir 1921) (за по-подробен поглед към критиките по адрес на методиката на Сапир вж. Schwegler 1990: 17).

Важна стъпка в посока разрешаване проблема с неясните определения представлява основополагащата работа на Джоузеф Грийнберг (Greenberg 1960), в която той предлага начин за квантитативно изследване на морфологично-типологичните аспекти на езика посредством индекси. В своята разработка Грийнберг използва методите за езикова класификация, предложени от Сапир, но с някои съществени модификации. За да избегне неясните определения, характерни за работата на Сапир, Грийнберг изчислява общо 10 индекса, всеки от които представлява съотношение между броя на определени граматични елементи (напр. префикси, суфикси, общия брой склоняеми морфемни елементи и др.) в дадена текстова проба – това формира числителя в уравнението – и общия брой думи в изследваната текстова проба или общата честота на употреба на даден друг граматичен елемент, напр. общия брой морфемни сглобки („morpheme junctures“ – Greenberg 1960: 185) – това формира знаменателя на уравнението. Така например индексът на синтетичността, изчислен от Грийнберг, представлява съотношението между общия брой на морфемите (M) в изследваната текстова проба и общия брой думи в нея (W), което е изразено в следната формула:  $M/W$  (срв. Greenberg 1960: 187). По този начин могат да се изчислят индекси за различни езици и лесно да бъдат сравнявани поради математическото си естество. Същевременно Грийнберг посочва и потенциала на предложения от него метод за изследване посоката на развитие на различните езици в исторически план (срв. Greenberg 1960: 194).

Предложената от Грийнберг методика вдъхновява редица лингвисти да приложат и усъвършенстват новия подход (срв. Krupa 1965, Krupa, Altman 1966, Slavičková 1968, Kelemen 1970, Altman, Lehfeldt 1973, Naarmann 1976), макар на малко по-късен етап методиката на Грийнберг да среща и някои значителни критики. Както ще видим и по-долу в настоящата разработка, един от основните проблеми на тази методология, който важи и за вдъхновените от нея подходи и остава не напълно разрешен и до днес, е проблемът с разграничението на думите в изследваната текстова проба и определянето на общия им брой (срв. Schwegler 1990: 20). Друг проблем откриваме в тенденцията езиците да проявяват различна степен на синтетичност в различните части от своя лексикон, т.е. даден език би могъл да бъде по-скоро аналитичен по отношение на един клас думи, а в друг клас думи да проявява повече синтетичност (срв. Schwegler 1990: 22). Въпреки тези слабости трябва да се отбележи, че работата на Грийнберг представлява важна стъпка към едно по-прецизно, внимателно и последователно изследване на морфологичната езикова типология.

Тук е редно да споменем и работата на българския езиковед Георги Герджиков (1983 и 1990), който също прилага математическия подход, характерен за Грийнберг. Различното в неговата работа е, че емпиричната база за анализ не е текстът, а самата парадигма на различните езикови подсистеми (глаголна, местоименна и т.н.). Този подход има предимството, че не е изложен до такава

степен на проблема с дефинирането на думата и разграничаването на отделните думи в свързан текст, тъй като при него не се анализира последователен текст. От друга страна обаче, разглеждането единствено на размера и сложността на парадигмите в даден език създава предпоставки за поне донякъде изкривени резултати. Нерядко в езиците съществуват граматически категории със свои специални форми и конструкции, които, макар и налични и познати на носителите на езика, се използват рядко и само в определени ситуации. Пример за такава категория откриваме в немския език, където употребата на преизказното наклонение *Konjunktiv I* е характерна основно за някои по-формални проявления на писмения език, като напр. журналистическите, правните или художествените текстове. Също в немски добре позната е частичната взаимна заменяемост на миналите времена *Präteritum* и *Perfekt*, които обаче се използват в различни езикови ситуации: *Perfekt* – в разговорния устен език, *Präteritum* – в писмените текстове. Изследването и анализът на автентични текстове, били те писмени или устни, дават възможност да се разгледа морфологията на даден език в контекста на реалната му употреба и да се получи една по-реалистична представа за нивото му на аналитичност и синтетичност. Това е и една от причините настоящата разработка да се фокусира върху този метод за квантитативно изследване на морфологичната езикова типология.

Прегледът на дотук споменатите публикации показва, че морфологичната езикова типология е един клон на лингвистиката, който е преживял чувствително развитие през 20. век, макар и за известно време в историята на тази наука да е останал на по-заден план. Същевременно обаче методиката на квантитативните морфологично-типологични изследвания все още е изправена пред някои проблеми, които възпрепятстват постигането на по-прецизни и надеждни резултати. В следващата част от настоящата работа фокусът ще бъде поставен върху съвременното състояние на методиката, както и на промените, които са настъпили в нея спрямо отминалите периоди на нейното развитие. На по-късен етап ще бъдат дискутирани и нейните проблеми, както и някои възможни решения.

## **2. Съвременни методи на квантитативната морфологична езикова типология**

В съвременния период подходът, предложен първоначално от Грийнберг, продължава да стои в основата на много изследвания в сферата на морфологичната типология. Между тези по-нови изследвания и оригиналната разработка на Грийнберг съществуват някои различия. Съвременните методи на квантитативната морфологична езикова типология се характеризират със следните особености.

2.1. Аналитичността и синтетичността на даден език или диалект са основните изследвани величини, при които в диахронен план се наблюдават различни акценти в инструментализирането на понятията *анализизъм* и *синте-*

*тизм*: в оригиналната разработка на Грийнберг индексът на синтетичността е само един от множество сходни индекси. Освен синтетизма посредством индекси той измерва и нивото на аглутинация, деривация, композиция, флексия, употреба на префикси и суфикси и други величини. В по-новите изследвания (срв. Szmrecsányi 2009, Szmrecsányi, Kortmann 2009 и 2011, Szmrecsányi 2016, Horsch 2021) се набляга основно на индекса на синтетичност, като обаче се изчислява и индекс на аналитичност. При това двете величини се изчисляват и разглеждат отделно.

2.2. Отдалечаване от идеята за чист аналитично-синтетичен континуум: още от момента на своето въвеждане от Аугуст фон Шлегел (срв. Schlegel 1818) понятията *аналитичен* и *синтетичен* се възприемат като пълни противоположности и не без основание, предвид това че аналитичните и синтетичните стратегии за кодиране на граматическа информация се различават значително една от друга. В съвременните изследвания обаче се откроява едно отдалечаване от възникналата междуременно идея, че тези две понятия се намират в двата края на един и същ континуум и че увеличението на аналитичността на един език задължително означава спад в синтетичността му и обратно (срв. Stankov 2024). По тази причина освен понятието *аналитично-синтетичен континуум* като инструмент за концептуализация и визуализация на различията между езиците по отношение на тяхната аналитичност и синтетичност, в съвременните изследвания (срв. Szmrecsányi, Kortmann 2011) се използва и понятието *типологично пространство* („typological space“) – диаграма, с чиято помощ различните езици или езикови разновидности могат да се сравняват нагледно чрез позиционирането им в двуизмерното пространство.

2.3. Стремеж към прецизиране и опростяване на работните дефиниции: както и в други области на съвременната наука, един от стремежите в по-новите изследвания в областта на морфологичната типология е формулирането на дефиниции за основните работни понятия (в случая това са понятията *аналитичен* и *синтетичен*), които са едновременно ясни, прецизни и кратки. Смерчаний и Кортман например работят със следните основни дефиниции: „overt *grammatical analyticity*, which we operationalize as the text frequency of free grammatical markers, and overt *grammatical syntheticity*, which we define as the text frequency of bound grammatical markers“ (Szmrecsányi, Kortmann 2011: 169). Тези две дефиниции все пак трябва да се разглеждат в контекста на съответното изследване, но изчистеността и яснотата на предложените определения са ясно различими.

2.4. Актуализирани формули за изчисление на индексите: начинът за изчисление на индекса на синтетичността е претърпял множество промени през годините, изминали от публикуването на разработката на Грийнберг до сега. Много от лингвистите, възприели подхода на Грийнберг, са правили промени по начина на изчисление на индексите, които той въвежда. Пример за актуална формула за пресмятане на индекса на синтетичността откриваме отново у

Смречаний и Кортман: „SYNTHETICITY INDEX =  $V/W \times 1,000$ “ (Szmrecsányi, Kortmann 2011: 174), където  $V$  представлява броя на свързаните граматически маркери (напр. окончания за спрежение на глаголи, склонение на прилагателни, число на съществителни и др.) в анализираната проба, а  $W$  обозначава общия брой на думите в пробата. Подобна е и формулата за изчисление на индекса на аналитичността: „ANALYTICITY INDEX =  $F/W \times 1,000$ “ (Szmrecsányi, Kortmann 2011: 174), където  $F$  представлява броя свободни граматични маркери (напр. спомагателни глаголи, съюзи, предлози и др.) в анализираната проба, а  $W$  отново е общият брой думи в пробата. За сравнение, както бе посочено и по-горе, формулата за индекс на синтетичността, която Грийнберг първоначално предлага, е  $M/W$ , където  $M$  е броят морфеми в пробата, а  $W$  е общият брой на думите в пробата (Greenberg 1960: 185).

Разделянето на индекса на синтетичността от индекса на аналитичността и тяхното отделно изчисление и разглеждане вероятно произтичат от концептуалното отдалечаване от идеята за стриктен аналитично-синтетичен континуум (вж. т. 2.2.). Същевременно това създава и предпоставките за изчислението на индекс на граматичността, който илюстрира общото количество граматическа информация, кодирано в текста. Този индекс от своя страна дава възможността да се изследва не само степента на свързаност или самостоятелност на граматическите маркери, но и на това как се развива количеството граматическа информация, която се изразява по явен начин в езика.

2.5. Употреба на методи, характерни за компютърната лингвистика: съвременните изследвания на морфологичната типология често използват електронни корпуси от текстове, които служат като източник на проби за анализ. Анализът на пробите може да бъде подпомогнат от различни видове софтуер като конкордансъри или програми за автоматичен морфологичен разбор (напр. AntConc, TreeTagger или TagAnt). Тези инструменти позволяват обработката на по-големи и повече на брой проби, което способства постигането на по-представителни резултати. Пробите, изследвани от Грийнберг (Greenberg 1960), например съдържат по 100 думи, докато тези, анализирани от Смречаний и Кортман (Szmrecsányi, Kortmann 2011), съдържат по 1000 думи. В допълнение в съвременните изследвания често се използват и тестове за статистическа значимост (напр. хи-квадрат тестове или Т-тест), които могат да дадат допълнително потвърждение за надеждността и реалността на получените резултати.

2.6. Употреба на понятието *граматичност*: в допълнение към индексите за аналитичност и синтетичност в някои съвременни изследвания се изчислява и индекс на граматичността, който представлява сбора на горните два индекса и отразява общото количество граматическа информация, кодирано в текста (за повече детайли относно понятието *граматичност* вж. Szmrecsányi 2009 и Szmrecsányi, Kortmann 2009).

Така представените специфики на съвременните изследвания позволяват открояването на някои предизвикателства пред изследователите, които ще бъдат систематизирани по-долу.

### 3. Някои проблеми пред съвременната квантитативна морфологична езикова типология

Видно е, че поне при част от съвременните изследвания на морфологичната езикова типология методиката, създадена от Грийнберг, все още е актуална, макар и в до известна степен видоизменена форма. Благодарение на постиженията на бързо развиващата се компютърна лингвистика подходът, въведен от Грийнберг, може да бъде използван при анализа на все по-големи текстови корпуси и да бъде допълнително усъвършенстван и обогатяван. Същевременно езиковедите, работещи в сферата на квантитативната морфологична езикова типология, са изправени и пред някои значителни методологически проблеми. Дискусията в настоящата разработка ще засегне четири от тях.

3.1. Липсата на универсална, ясна и изчерпателна дефиниция за понятието *дума*: както за методиката, предложена от Грийнберг, така и за описаните в последния раздел съвременни подходи за изчисление на индекси на синтетичността и аналитичността понятието *дума* е от ключово значение, тъй като присъства във всяка от представените формули. Изчисляването на индексите е невъзможно, без да знаем с точност броя думи в анализираната проба. Но какво всъщност е думата и едно и също нещо ли е тя в български, английски, немски, турски и в другите езици по света? За да могат така изчислените индекси да бъдат сравними и да бъдат показателни за разликите между морфологичните системи на различните езици, е нужно понятието *дума* да бъде дефинирано ясно и изчерпателно. В най-добрия случай една такава дефиниция би била и универсална, т.е. приложима към всички езици по света. Без съмнение, през вековете понятието *дума* е било дискутирано многократно, но според Шwegler (Schwegler 1990: 29) до настоящия момент не е била предложена дефиниция, която да работи за всеки език.

3.2. Твърде общата употреба на понятията *аналитичен* и *синтетичен*: в наши дни тези две понятия са придобили популярност както сред професионалните лингвисти, така и сред обикновените любители на езиците. Те редовно се употребяват при описанието на цели езикови системи. Например английският нерядко бива определян като силно аналитичен език със синтетично минало, докато латинският се представя като типичен пример за синтетичен език. Шwegler също отбелязва този проблем, както и някои потенциални решения, и посочва: “they must be understood not as quantifiable absolutes but as the *rough measure of the overall morphemic interdependency of speech units*” (Schwegler 1990: XV). Същността на проблема тук е, че понятията *аналитичен* и *синтетичен* са сравнително груби инструменти за езикова типологизация и

приложението им към цели езици не води до особено точни, задълбочени или нюансирани изводи относно структурата и функционирането на тези езици.

3.3. Неясният статус на някои класове думи, считани за свободни граматични маркери: както бе споменато по-горе, числителят на формулата за изчисление на индекса на аналитичността в някои съвременни изследвания (напр. Szmrecsányi, Kortmann 2009, Szmrecsányi, Kortmann 2011, Horsch 2021, Stankov 2024) представлява броят свободни граматични маркери в анализиранията проба. Според Сзмречаний и Кортман в английския към тези маркери принадлежат следните класове думи: „conjunctions, subjunctions, and prepositions [...]; determiners, articles, and *wh*-words [...]; existential *there* [...]; pronouns [...]; the tokens *more* and *most*; the infinitive marker *to* [...]; modals [...]; the negator *not* [...], auxiliary *be* [...], auxiliary *do* [...] and auxiliary *have*“ (Szmrecsányi, Kortmann 2011: 174). Предлозите и съюзите обаче се възприемат от някои лингвисти като класове думи, които имат не само граматическо, но и лексикално значение. Затова на дневен ред стои въпросът: редно ли е тези класове думи да се считат за свободни граматични маркери и да участват при изчисляването на индекса на аналитичността и ако да, на какво основание?

3.4. Потенциалът за изкривени резултати при сравнение нивата на синтетичност на аглутинативни и флективни езици: описаната тук методология за изчисляване на индекси на аналитичност и синтетичност работи добре при сравнението на флективни езици от индоевропейското езиково семейство, но приложението ѝ за изследване на други типове езици и най-вече за тяхното сравнение с флективните езици би могло да доведе до изкривени резултати. Проблемът произтича основно от факта, че за аглутинативните езици е характерно върху корена на думата да се натрупват наставки, всяка от които представлява свързан граматичен маркер. При аглутинативните езици обикновено всяка наставка изразява по една граматическа категория, поради което често се получава наслагване на наставки една след друга. При флективните езици обаче свързаните морфеми с граматическа функция понякога изразяват повече от една граматическа категория.

В настоящия си вид методите на квантитативната морфологична типология за изчисление на степента на аналитичност и синтетичност на езиците не вземат предвид този факт. В резултат на това анализът на проби от аглутиниращи и от флективни езици вероятно би показал по-висока степен на синтетичност у аглутиниращите езици. Съчетанието на повече от едно граматическо значение в рамките на една морфема, характерно за флективните езици, обаче е олицетворение на синтеза на граматическа информация. Необходимо е по някакъв начин този синтез да бъде отразен в начина на изчисление на индекса на синтетичността, за да може сравнението с аглутинативните езици, които синтезират чрез наслагване на наставки, да бъде по-реалистично.

#### 4. Потенциални решения

Гореизброените проблеми със сигурност не изчерпват критичните моменти, свързани с методологията на квантитативната морфологична езикова типология. В дискусиата по-долу ще бъдат дадени някои идеи за разрешаването или поне за смекчаването на споменатите проблеми.

4.1. Употреба на ортографичната дума: проблемът с дефинирането на думата е може би най-старият от изброените тук проблеми. Правени са много опити за създаване на универсална дефиниция, приложима за всички езици, но в световен мащаб езиковото разнообразие по отношение на думата изглежда е твърде голямо, за да може да бъде обхванато от една универсална дефиниция (за обзор на различните подходи към тази задача вж. Schwegler 1990, който има цяла глава по темата). Настоящата разработка няма за цел да предложи нов кандидат за такава универсална дефиниция, а по-скоро да представи аргументи за това защо ортографичната дума може да се използва като емпирична единица в квантитативните морфологични типологически изследвания, макар и невинаги да отговоря съвършено на това, което носителят на един език подсъзнателно възприема като дума.

На първо място, важно е да се обърне внимание на факта, че понятията *аналитичен* и *синтетичен* представляват сравнително широкомащабни и не-прецизни инструменти за категоризиране на езиковото разнообразие. Следователно и класификацията на езици чрез тях е по-обща и широкообхватна. По тази причина квантитативният анализ на аналитичността и синтетичността на един език може да бъде информативен и полезен за междуезиково и диахронно сравнение, дори когато за целта се използва стандартната ортографична дума. Изследването на Грийнберг (Greenberg 1960), както и разработките на Смерчаний, Кортман и Хорш (Szmrecsányi, Kortmann 2009, Szmrecsányi, Kortmann 2011, Horsch 2021) са добри свидетелства за ползата от такива сравнителни анализи. От друга страна, по-детайлната и прецизна езикова класификация така или иначе изисква разглеждането на отделните езикови подсистеми, както и употребата на повече специализирани понятия.

На следващо място, в много от съвременните езици, които са били кодифицирани и разполагат със своя писменост, дори тя да е базирана на чужда ортографична система, разделението на думите една от друга често следва естествената интуиция на носителите на езика, която може да бъде добър показател за това какво представлява една дума в дадения език. Разбира се, тук могат да се дадат контрапримери, тъй като вероятно в почти всеки език има думи, за които дори самите носители на езика не са сигурни дали представляват една словоформа, или две, или повече отделни такива. В допълнение с течение на времето това отчасти подсъзнателно възприятие за единството на една сложна словоформа може да се променя: „например сложния предлог *помежду* сега смятаме за една дума, но по-рано той се е писал разложено *по между* като две

думи“ (Стоянов 1993). Все пак обаче в мнозинството от случаите носителите на даден език могат сравнително точно да определят думите, от които се състои едно писмено или устно изказване. Ако приемем, че писмената норма в голям брой езици се базира върху тази езикова интуиция на носителите на езика, можем да очакваме, че ортографичната дума е достатъчно надеждна емпирична единица за изчислението на индекси на аналитичност и синтетичност.

В този ред на мисли трябва да бъде споменат добре известният факт, че писмената норма бавно възприема езиковите промени, които настъпват в устната реч. Съответно резултатите, получени при употребата на ортографичната дума за изчислението на индекси на аналитичност и синтетичност, не отразяват най-актуалното състояние на езика. Все пак обаче сравнението на получените по този начин индекси за различни езици може да бъде информативно, защото забавеното развитие на ортографията спрямо устния език е закономерност, която вероятно важи за по-голямата част от езиците по света.

В полза на употребата на ортографичната дума говори и още един аргумент, който е по-скоро прагматичен. Анализът на пробите не изисква допълнителна обработка на текста, за да бъдат изчислени индексите на аналитичност и синтетичност. Ако бъде възприета по-нестандартна дефиниция за дума, която значително се отклонява от стандартната ортографична дума, анализаторът на пробите би трябвало да обработи всяка проба ръчно, така че думите в текста да отговарят на новата дефиниция. Това би било необходимо, освен ако не е наличен специализиран софтуер, който би могъл да автоматизира процеса. Създаването на такъв софтуер обаче изисква по-значителни финансови или времеви ресурси, защото методиката, използваща нестандартна дефиниция за дума, би била иновативна, а не всеки лингвист има нужните умения, за да създаде сам софтуера, който би могъл да извърши обработката на текста автоматично.

Ръчната обработка на текста обикновено отнема доста време, което означава, че за да бъде изследването актуално, то не би могло да се базира върху множество проби, тъй като обработката им би отнела твърде много време. От друга страна обаче, изследване, базирано върху малко количество текст, е по-малко представително за езиковата система. Оттук следва, че в отсъствието на софтуер, който би могъл да трансформира автоматично един автентичен текст в текст, използващ думи, различни от стандартните за езика, ортографичната дума остава добра алтернатива за постигане на добро ниво на представителност на изследването.

4.2. Самостоятелен морфологично-типологичен анализ на отделните подсистеми на езика: за да се подобри полезността на понятията *аналитичен* и *синтетичен* за целите на езиковата типология, те биха могли да се прилагат към отделните подсистеми на езика. Например вместо един език да се характеризира типологически като едно цяло чрез изчислението на индекси за

аналитичност, синтетичност и граматичност, тези изчисления биха могли да се направят поотделно за неговата глаголна, именна и местоименна система, както и за другите му морфологични подсистеми. По този начин биха могли да се направят по-прецизни и полезни заключения и междуезикови сравнения. За целта обаче настоящата методология ще трябва да бъде модифицирана, за да може да се приложи селективно върху отделните подсистеми на езика.

4.3. Критерият *синсемантичност* като инструмент за прецизиране подбора на части на речта, считани за свободни граматични маркери: лексикалната стойност на някои от свободните елементи, които носят и граматическо значение (напр. предлози и съюзи), не може да се отрече. Въпреки това в доста съвременни изследвания в областта на морфологичната типология те се считат за свободни граматични маркери без оглед на лексикалната им стойност и се включват в изчислението на индекса на аналитичността. Смречаний прави това на основание тяхната синсемантичност (Szmrecsányi 2016: 95). Като синсемантични елементи тези лексеми могат да реализират лексикалното си значение само в комбинация с и в зависимост от други лексеми. Същевременно обаче те често изразяват взаимоотношения между конституентите на изречението, които биха могли да бъдат изразени по синтетичен начин. Срв. напр. в български: *Бащиният ми дом* и *домът на баща ми*. Във втория пример притежанието е изразено чрез предлог, докато в първия това взаимоотношение е изразено чрез прилагателното име. В съвременния български език и двата варианта са възможни. Също и в немски: *das Haus meines Vaters* и *das Haus von meinem Vater*. И тук във втория пример предлогът показва притежателното взаимоотношение, докато в първия то е изразено по синтетичен начин посредством родителен падеж. Тези примери демонстрират, че предлозите често играят функционална роля и могат да бъдат считани за маркери, характеризиращи аналитичния езиков строй.

Друго възможно решение за този проблем, в случай че синсемантичното естество на предлозите и съюзите се окаже недостатъчно основание за тяхната класификация като свободни граматични маркери, е те да се разглеждат поотделно при анализа на текстови проби, в съответния си контекст и да се определя на индивидуален принцип дали даден елемент носи граматична информация, или има основно лексикално значение. В зависимост от това ще се определя и дали даденият елемент да участва в изчислението на индекса на аналитичността на съответната проба. Този подход изисква по-задълбочен анализ и съответно е по-трудоемък и бавен. Същевременно трябва да се определят и ясни критерии, според които анализаторът да може последователно да преценява кога един предлог или съюз е използван като лексикален елемент и кога функцията му е граматическа.

4.4. Модификатор на синтетичната тежест като инструмент за получаване на адекватни резултати при сравнението на синтетичността на флективни

и аглутинативни езици: различното естество на аглутинативните и флективните езици представлява предизвикателство за методологията за измерване на аналитичността и синтетичността, използвана от Сзмречаний, Кортман и Хорш (Szmrecsányi, Kortmann 2009: 173, Horsch 2021: 346), тъй като аглутинативните езици имат потенциала да дадат много по-високи резултати за синтетичност от флективните езици, а именно флективните езици са прочути със своята синтетичност, особено някои древни езици, напр. класическият латински или санскрит.

Проблемът на актуалната методология е, че тя не е сензитивна за броя граматически категории, които един свързан граматичен маркер изразява едновременно. Този недостатък би могъл да се коригира чрез въвеждането на модификатор на тежестта, който да увеличава синтетичната тежест на маркерите, изразяващи повече от една граматична категория. Така те биха увеличили повече числителя на формулата за изчисление на индекса на синтетичността (за повече информация относно формулата вж. т. 2.4. горе). Например наставката *-те* в *книгите* би могла да има стойност 2 вместо 1 при броенето на свързаните граматични маркери, защото кодира едновременно категорията *определеност* и категорията *число*. В същото време морфемата *-s* в *the books* би имала стойност 1, защото кодира само една граматична категория – *число*. Разбира се, за да бъде тази модификация продуктивна, ще бъде нужно да се даде ясно определение на това какво е граматична категория и какви граматични категории съществуват и могат да се изразяват по синтетичен начин в изследвания език.

### **5. Заключение – възможности и предизвикателства пред квантитативната морфологична езикова типология**

В настоящата разработка бяха разгледани четири проблема на методологията, прилагана в някои съвременни изследвания в сферата на квантитативната морфологична езикова типология. В обобщен вид става дума за проблеми, засягащи универсалността, обхвата и приложимостта на някои ключови понятия, използвани в разработките, като напр. *дума*, *аналитичен*, *синтетичен*, *свободен граматичен маркер*. В дискусиата по-горе бяха предложени конкретни решения за някои от описаните проблеми, но за липсата на универсална и изчерпателна дефиниция за понятието *дума* истински задоволително решение все още не е постигнато, макар и да съществуват частични ситуативни решения, като напр. употребата на ортографичната дума в типологичните изследвания на морфологичната аналитичност и синтетичност.

Успешното разрешаване на дискутираните по-горе проблеми би довело до по-представителни и надеждни резултати от квантитативните морфологично-типологични изследвания. Междувременно съвременното крие немало възможности за това поле на лингвистиката. Благодарение на постиженията на компютърната лингвистика и с помощта на глобалната свързаност посред-

ством интернет езиковите типолози имат на разположение все повече и по-богати на информация корпуси, които могат да служат като източник на материал за профилирани квантитативни типологични изследвания на морфологията на езиците по света. Специализираният софтуер и бързо развиващите се големи езикови модели биха могли също да допринесат много за скоростта и обхвата на анализа в такива разработки. Тук можем да споменем и разрастващата се езикова документация, която дава достъп на изследователите до материал от все повече екзотични и не така добре изследвани досега езици. Благодарение на тези и други съвременни възможности квантитативната езикова типология би могла да остане релевантна в сферата на лингвистиката в една епоха на динамичен и широкообхватен научно-технически прогрес.

### БИБЛИОГРАФИЯ

**Братанова 2021:** Братанова, Б. Иновационни процеси в съвременната граматика на английския език (с оглед на техните възможни паралели в българския език). – В: *Български език*, Приложение, 68 (2021), 126 – 148. // **Bratanova 2021:** Bratanova, B. Inovatsionni protsesi v savremennata gramatika na angliyskiya ezik (s ogleed na tehните vazmozni paraleli v balgarskiya ezik). – In: *Balgarski ezik*, Prilozhenie, 68 (2021), 126 – 148 <[https://www.balgarskiezik.eu/p-2021/8\\_B\\_BRATANOVA\\_str\\_126-148\\_BG.pdf](https://www.balgarskiezik.eu/p-2021/8_B_BRATANOVA_str_126-148_BG.pdf)> [01.06.2024].

**Герджиков 1983:** Герджиков, Г. Тенденцията към аналитизъм – определение, метод за измерване, причини, следствия. – В: *Съпоставително езикознание*, т. 8, кн. 5, 46 – 55. // **Gerdzhikov 1983:** Gerdzhikov, G. Tendentsiya kam analitizam – opredelenie, metod za izmervane, prichini, sledstviya. – In: *Sapostavitelno ezikoznanie*, vol. 8, Nr. 5, 46 – 55.

**Герджиков 1990:** Герджиков, Г. Развоят на индоевропейските езици към аналитизъм и някои всеобщи типологически зависимости. – В: *Съпоставително езикознание*, т. 15, кн. 4 – 5, 151 – 157. // **Gerdzhikov 1990:** Gerdzhikov, G. Razvoyat na indoeuropeyskite ezitsi kam analitizam i nyakoi vseobshti tipologicheski zavisimosti. – In: *Sapostavitelno ezikoznanie*, vol. 15, Nr. 4 – 5, 151 – 157.

**Петров 2021:** Петров, В. Иновации в граматическите системи на съвременния български език и на новогръцкия език. – В: *Български език*, Приложение, 68 (2021), 149 – 164. // **Petrov 2021:** Petrov, V. Inovatsii v gramaticheskite sistemi na savremenniya balgarski ezik i na novogratskiya ezik. – In: *Balgarski ezik*, Prilozhenie, 68 (2021), 149 – 164 <[https://www.balgarskiezik.eu/p-2021/9\\_V\\_PETROV\\_str\\_149-164\\_BG.pdf](https://www.balgarskiezik.eu/p-2021/9_V_PETROV_str_149-164_BG.pdf)> [01.06.2024].

**Стоянов 1993:** Стоянов, С. Дума и нейното значение. – В: Тилков, Д., С. Стоянов, К. Попов (ред.). *Грамматика на съвременния български книжовен език. Том 2. Морфология*. София: Издателство на БАН, 10 – 23. // **Stoyanov 1993:** Stoyanov, S. Duma i neynoto znachenie. – In: Tilkov, D., S. Stoyanov, K. Popov (red.). *Gramatika na savremenniya balgarski knizhoven ezik. Tom 2. Morfologiya*. Sofia: Izdatelstvo na BAN, 10 – 23.

**Altman, Lehfeldt 1973:** Altman, G., W. Lehfeldt. *Allgemeine Sprachtypologie. Prinzipien und Messverfahren*. Munich: Fink.

**Greenberg 1960:** Greenberg, J. A Quantitative Approach to the Morphological Typology of Language. – In: *International Journal of American Linguistics*, vol. 26, No. 3, 178 – 194 <[https://archive.org/details/sim\\_international-journal-of-american-linguistics\\_1960-07\\_26\\_3/page/178/mode/2up](https://archive.org/details/sim_international-journal-of-american-linguistics_1960-07_26_3/page/178/mode/2up)> [01.07.2024].

**Haarmann 1976:** Haarmann, H. *Grundzüge der Sprachtypologie. Methodik, Empirie und Systematik der Sprachen Europas*. Stuttgart: Kohlhammer.

**Horsch 2021:** Horsch, J. Typological Profiling of English, Spanish, German and Slovak: A Corpus-based Approach. – In: *Journal of Linguistics/Jazykovedný časopis*, vol. 72 No. 2, 342 – 352 <[doi: 10.2478/jazcas-2021-0032](https://doi.org/10.2478/jazcas-2021-0032)> [01.07.2024].

**Kelemen 1970:** Kelemen, J. Sprachtypologie und Sprachstatistik. – In: Deszö, L., P. Hajdú (ed.). *Theoretical Problems of Typology and the Northern Eurasian Languages*. Amsterdam: Grüner, 53 – 63.

**Krupa 1965:** Krupa, V. On Quantification of Typology. – In: *Linguistics*, 12, 31 – 36.

**Krupa, Altman 1966:** Krupa, V., G. Altman. Relations Between Typological Indices. – In: *Linguistics*, 24, 29 – 37.

**Sapir 1921:** Sapir, E. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace & World Inc.

**Schlegel 1808:** Schlegel, F. W. *Über die Sprache und Weisheit der Indier: ein Beitrag zur Begründung der Alterthumskunde: nebst metrischen Übersetzungen indischer Gedichte*. Heidelberg, Deutschland: Mohr und Zimmer.

**Schlegel 1818:** Schlegel, A. W. *Observations sur la langue et la littérature provençales*. Paris: Librairie grecque-latine-allemande <<https://gallica.bnf.fr/ark:/12148/bpt6k63239579/f24.item.texteImage#>> [01.07.2024].

**Schwegler 1990:** Schwegler, A. *Analyticity and Syntheticity. A Diachronic Perspective with Special Reference to Romance Languages*. Mouton de Gruyter. Berlin/New York.

**Slavičková 1968:** Slavičková, E. Towards a Typological Evolution of Related Languages. – In: *Travaux Linguistiques de Prague*, 3. University, AL: University of Alabama Press.

**Stankov 2024:** Stankov, N. Development Trends in the Morphological-typological Profile of German. – In: *Studia Philologica Universitatis Velikotarnovensis*, vol. 43, No. 1, 37 – 49.

**Szmrecsányi 2009:** Szmrecsányi, B. Typological Parameters of Intra-lingual Variability: Grammatical Analyticity versus Syntheticity in Varieties of English. – In: *Language Variation and Change*, 21. Cambridge: Cambridge University Press, 319 – 353.

**Szmrecsányi, Kortmann 2009:** Szmrecsányi B., B. Kortmann. Between Simplification and Complexification: Non-standard Varieties of English around the World. – In: Sampson, Gil, Trudgill (eds.). *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 64 – 79.

**Szmrecsányi, Kortmann 2011:** Szmrecsányi B., B. Kortmann. Typological Profiling: Learner Englishes versus Indigenized L2 Varieties of English. – In: Mukherjee, J., M. Hundt (eds.): *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam/Philadelphia: John Benjamins Publishing Co., 167 – 188.

**Szmrecsányi 2016:** Szmrecsányi, B. An Analytic-synthetic Spiral in the History of English. – In: Van Gelderen, E. (ed.): *Cyclical Change Continued*, 93 – 112 <[doi:10.1075/la.227](https://doi.org/10.1075/la.227)> [01.07.2024].