

МЕТОДОЛОГІЯ СОЦІОЛІНГВІСТИЧНОГО АНАЛІЗУ ІСТОРИЧНИХ МОВНИХ КОРПУСІВ

У статті розглядається методологія аналізу даних, вилучених з історичних мовних корпусів. Визначається потенціал широкого використання кількісних аналітичних методів у соціолінгвістичних діахронічних дослідженнях варіативності й мовних змін. Досліджуються аналітичні інструменти, що забезпечують обробку комплексних даних та дозволяють проводити детальний аналіз мовних подій.

Ключові слова: історична соціолінгвістика, корпусна лінгвістика, кількісні і якісні методи аналізу даних, принцип репрезентативності

Семененко Г.Н. Методология социолингвистического анализа исторических языковых корпусов. В статье рассматривается методология анализа данных, извлеченных из исторических языковых корпусов. Определяется потенциал широкого использования количественных аналитических методов в социолингвистических диахронических исследованиях вариативности и языковых изменений. Исследуются аналитические инструменты, обеспечивающие обработку комплексных данных, которые позволяют проводить детальный анализ языковых событий.

Ключевые слова: историческая социолингвистика, корпусная лингвистика, количественные и качественные методы анализа данных, принцип репрезентативности

Semenenko G.M. Methodology for sociolinguistic analysis of historical language corpora. This paper discusses the methodology of analyzing data extracted from historical language corpora. The potential of the large-scale application of quantitative analytical methods in diachronic sociolinguistic studies of language variation and change is determined, as well as the role of these analytical instruments in the comprehensive data processing and detailed analysis of speech events.

Keywords: historical sociolinguistics, corpus linguistics, quantitative and qualitative methods of data analysis, the principle of representativity

Основними рушійними силами у формуванні соціолінгвістики як окремого напрямку лінгвістичних досліджень у 60-ті роки ХХ ст. прийнято вважати невдоволення низки дослідників попередніми науковими парадигмами, нове визначення та переформатування традиційної діалектології, підвищений інтерес лінгвістів до соціології та її можливостей, і так звана “кількісна революція” [Hernández-Campoy, Schilling 2012]. Останній фактор знайшов своє вираження у спробах адаптувати методи точних наук до потреб мовознавства, розробити кількісну лінгвістичну парадигму, в якій соціальні, культурні та контекстуальні фактори можуть використовуватися для опису, пояснення та передбачення моделей мовного варіювання і змін. Комп’ютеризовані процедури пропонують новий **актуальний** підхід до збору і обробки даних дослідження, часто зосереджуючи увагу науковців на фактах, які могли бути проігнорованими або непоміченими при традиційному аналізі. Відповідно **метою** статті є визначення і характеристика кількісних і якісних методів аналізу даних, отриманих в результаті обробки історичних мовних корпусів.

Принципи репрезентативності і узагальненості з самого початку були фундаментальними в методології варіативних соціолінгвістичних досліджень. Вони й досі вважаються ідеалом, якого потрібно прагнути, хоч соціолінгвісти давно визнали, що досягнення абсолютної статистичної репрезентативності є надзвичайно важким завданням, яке, до того ж, не завжди є обов’язковим для отримання ґрунтовних результатів [Sankoff, Labov 1979, с. 51–52]. Принцип репрезентативності передбачає, що всі члени мовної спільноти повинні мати рівні шанси бути обраними типовими інформантами, крім того, інформанти повинні обиратися з урахуванням соціальних і демографічних особливостей населення в цілому. Принцип узагальненості наголошує на тому, що будь-яке соціолінгвістичне дослідження повинно відповідати двом критеріям для того, щоб отримані результати могли вважатися точними і екстраполюватися на мовну поведінку всього населення, або мовної спільноти: ймовірності (ті ж самі результати можуть бути досягнуті при повторному дослідженні того самого явища) і суб’єктивності (два незалежних один від одного дослідника, вивчаючи одне і те саме явище, повинні прийти до однакових результатів) [Bailey, Tillery 2004, с. 1].

Історичну соціолінгвістику часто критикують за недостатню репрезентативність і валідність результатів. На жаль, цього неможливо уникнути, оскільки соціолінгвістичне дослідження історичних етапів розвитку будь-якої мови ґрунтується на мовних пам'ятках відповідних періодів, які у той чи інший спосіб є неповними або нерепрезентативними; крім того, джерелом інформації також є соціокультурна ситуація досліджуваного періоду, яка дуже рідко підлягає прямому спостереженню і аналізу дослідників. Основними проблемами соціолінгвістичних діахронічних досліджень є: репрезентативність, емпірична валідність, інваріантність, автентичність, авторство, соціальна й історична валідність й ідеологія стандарту [Hernández-Campoy, Schilling 2012]. Однак, теоретичні та процедурні проблеми присутні і в, на перший погляд, точно вивірених методології варіативної соціолінгвістики, що відображається в нестабільній відповідності досліджень принципу узагальненості. Без ретельного дослідження впливу методології на результати дослідження неможливо точно стверджувати, наскільки узагальненими, тобто репрезентативними і надійними є такі дослідження [Bailey, Tillery 2004, с. 13]. Тому важко вимагати від історичної соціолінгвістики відповідності тим високим стандартам, які для себе встановила соціолінгвістика.

Оскільки базою даних для історичного соціолінгвістичного наукового пошуку служать письмові тексти різних жанрів, що збереглися від самих ранніх етапів розвитку мови, важко переоцінити роль комп'ютеризованих мовних корпусів у полегшенні дослідження. Кількість мовних корпусів, доступних для досліджень соціолінгвістів, є ознакою плідних взаємин лінгвістики, історії й інформаційних технологій та їхнього активного поглиблення в останні десятиліття. Ресурси на зразок Гельсінського корпусу, низки мовних корпусів групи Brown, репрезентативного історичного корпусу різножанрових англійських текстів (ARCHER) і корпусу ранньоанглійського листування (PCEEC) сприяли активним дослідженням лексичних і граматичних змін в довгостроковій та короткостроковій перспективах. Історичні мовні корпуси надають змогу досліджувати варіативність і мовні зміни по жанрам, регістрам та діалектам. Існуючі діахронічні корпуси відрізняються часовим охопленням, проте основний принцип укладання таких корпусів полягає в незмінності типу тексту, що уможливорює дослідження процесу мовних змін в текстах одного жанру з урахуванням якнайменшої кількості перешкод.

Незважаючи на те, що у сучасному мовознавстві корпусна лінгвістика асоціюється перш за все із форматом, придатним для комп'ютерного аналізу, мовні корпуси уклалися задовго до початку комп'ютерної ери такими видатними дослідниками, як, наприклад, Самуель Джонсон, Александр Дж. Елліс, Джозеф Райт, Джеймс Маррі, Гаролд Ортон та Рандольф Квірк для проведення граматичних, лексикографічних і діалектологічних досліджень [Francis 1992]. Так, працюючи над своїм відомим "Словником англійської мови", виданим у 1755 р., Самуель Джонсон зібрав великий, рукописний корпус ілюстративного матеріалу з літературних, філософських та наукових джерел (приблизно 150000 прикладів для 40000 тис слів) з метою фіксації особливостей англійської лексики. Е.Сепір, Л. Блумфілд та багато інших дослідників наголошували на необхідності використання широкої емпіричної бази у дослідженнях [McEnergy, Xiao, Tono 2006]. І хоча ранні корпуси склалися з паперових карток, робота по збору і класифікації ілюстративного матеріалу проводилася так само, як і сучасний якісний корпусний аналіз. Виникнення і становлення корпусної лінгвістики як окремої методології дослідження у сучасному електронному форматі асоціюється з новаторською роботою У. Нельсона Френсіса і Генрі Кучери по розбудові корпусу Brown в 1960-х роках ХХ ст. Мовний корпус, який складався з одного мільйона слів, започаткував кількісні корпусні дослідження і створив підґрунтя для розбудови методів, що лежать в основі сучасного якісного корпусного аналізу.

Якісний корпусний аналіз – це методологія досліджень мовних явищ в контексті автентичних комунікативних ситуацій, що збережені в електронному форматі у вигляді мовних корпусів і підготовлені до аналізу за допомогою сучасних технологій. Дослідники, які використовують якісний корпусний аналіз як методологічну базу, використовують

індуктивний підхід до емпіричних досліджень взаємодії форми і функцій мовних одиниць з різними зовнішніми чинниками у процесі спілкування (вік, гендер, освіта, соціально-економічний стан; місце і час комунікативної події; стосунки між співрозмовниками; мовленнєва модальність і т.д.). Методологічний симбіоз якісного корпусного аналізу з “традиційними” якісними підходами є органічним, оскільки всі вони орієнтовані на аналіз мовних явищ через пояснення різноманіття контекстуальних факторів, які розміщують результати якісних досліджень в екосистемі людського спілкування. Цей симбіоз є взаємовигідним, оскільки без досягнень традиційної якісної методології в міждисциплінарних дослідженнях останніх ста років, корпусна методологія була б позбавлена відточеності методологічного апарату для опису і інтерпретації комплексної природи людської комунікації.

Незважаючи на те, що якісний корпусний аналіз тлумачиться як такий, що не має справу з частотністю і статистичною класифікацією лінгвістичних рис, значення комбінованого підходу до аналізу мовних явищ не викликає ніяких сумнівів. Висновки, отримані в процесі кількісного корпусного аналізу, можуть бути розвинуті, пояснені і проілюстровані, використовуючи інтерпретаційну спроможність якісної методології, що зумовлює паралельне використання обох підходів.

Будь-яке дослідження тільки виграє від використання кількісних методів, оскільки якісні методи часто критикують за трудність у науковому відтворенні та підтвердженні результатів аналізу. Кількісні корпусні дослідження організовані таким чином, щоб дані щодо частотності вживання могли дати відповіді на поставлені питання. Важливе значення також мають так звані експериментальні технології, створені для трансформації складних рядів даних в підсумкову (часто візуалізовану) репрезентацію, що пізніше інтерпретується аналітиками і часто призводить до формулювання нової гіпотези.

Корпусні соціолінгвістичні дослідження присвячені порівнянню частотності об'єктивно обчислюваних мовних ознак, іншими словами, кількісному аналізу отриманих прикладів. Типове завдання дослідження в рамках варіативної соціолінгвістики формулюється так: чи вірно, що змінна x використовується по-різному в корпусі А і в корпусі Б (наприклад, корпуси текстів різних періодів, різних регіональних/національних варіантів, усні або писемні, тексти різних жанрів)? Для отримання відповіді це запитання необхідно трансформувати у гіпотезу – твердження, що підлягає перевірці. В залежності від результатів тестів гіпотеза приймається або відкидається. Перевірка гіпотези – отримання статистичного висновку на основі аналізу даних – передбачає використання кількісних методів для встановлення ймовірності того, чи є гіпотеза вірною [Cantos 2012, с. 104].

Кількісні дослідження змін у мовній структурі в різні проміжки часу допомагають у вирішенні глобальних лінгвістичних проблем, наприклад, встановлення часу та особливостей процесу перебігу мовної зміни, виділення окремих фаз такого процесу, визначення взаємопов'язаності або, навпаки, незалежності змін у формальних та функціональних ознаках мовної форми, виокремлення факторів та рушійних сил процесу мовних змін, їх відносну важливість у просторі і часі, порівнянні випадків мовних змін в минулому з мовними змінами, що відбуваються у сьогоденні. Не викликає заперечень, що кількісні дослідження діакронії мають багато спільного із синхронними кількісними студіями [Tagliamonte 2006] на теоретичному і методологічному рівнях. Така схожість не є випадковою, оскільки мовна варіативність – це ключовий фактор до пояснення мовних змін.

Історична лінгвістика залежить від аналізу автентичних даних, проте не всі проблеми дослідження змушують вчених генерувати кількісні дані. Низка процесів мовних змін виявляє себе в якісних відмінностях, так, лексичні одиниці виходять з ужитку, або моделі порядку слів, які колись домінували, стають периферійними. Однак, ретельніша цифрова обробка діакронічних даних стає необхідною у випадку не простої констатації факту зміни, а під час дослідження внутрішньої динаміки такої зміни. Кількісні дослідження є невід'ємною частиною багатофакторних пояснень процесу мовних змін, що включають формальні, функціональні, зовнішні соціальні чинники, які необхідно взяти до уваги; вони значно

полегшують аналіз взаємодії між формальними, функціональними та зовнішніми чинниками для визначення того, чи має кожен окремий чинник той самий вплив незалежно від інших чинників [Hilpert, Gries 2014]. Кількісні методи використовуються для дослідження взаємодії часу з одного боку і формальних, функціональних та зовнішніх соціальних чинників з іншого боку з метою визначення дії одного і того самого чинника в кожному часовому проміжку, або динаміки впливу конкретного фактора на вибір мовців. Без аналізу такої взаємодії неможливо робити порівняння мовних особливостей в різні періоди розвитку.

У низці досліджень використовуються комплексні підходи, часто у формі багатофакторного регресійного аналізу. Регресійний аналіз має на меті передбачення результату залежної змінної (або реакції) на основі однієї або більше незалежних змінних (або предикторів). Важливо, що і реакція, і предиктори можуть бути різного типу, наприклад, бінарними (синтетичні і аналітичні форми ступенів порівняння прикметників), категоріальними або порядковими (істоти / неістоти/ абстрактні), або цифровими (тривалість звучання або довжина слова в фонемах); з урахуванням характеру залежної змінної використовується бінарна логістична регресія, поліноміальна, порядкова логістична регресія, або лінійна регресія. Основною перевагою регресійних моделей є те, що вони дозволяють дослідникам розглядати ефект декількох предикторів (та їх взаємодію) одночасно [Ваауен 2008, с. 165–236] і визначати, які з предикторів мають значний вплив на мовний вибір та як саме реалізується цей вплив (чи робить конкретний предиктор вибір більш ймовірним). Проте, незважаючи на великі переваги регресійного моделювання, є дослідження, які не передбачають порівняння частотності певних явищ в різні історичні періоди [Gries 2011]. Наприклад, досліджуючи частоту вживання абсолютних дієприкметникових речень в середньо- і ранньомовноанглійському періодах, недостатньо простого аналізу окремих прикладів, або таблиць частоти вживання абсолютних речень у порівнянні з іншими фінітними і нефінітними синтаксичними конструкціями. Необхідно, перш за все, ідентифікувати контексти, в яких вживаються ці речення, а потім визначити формальні, функціональні і зовнішні соціальні особливості цих контекстів. На абстрактному рівні, відповіді, які шукає дослідник, ґрунтуються на трьох критеріях: необхідно визначити, що “в часовий період X контекстуальна особливість Y сприяла вибору мовців на користь нової синтаксичної конструкції з відносною силою Z” [Hilpert 2013, с. 50]. Аналізуючи вплив низки контекстуальних особливостей в певні часові проміжки, дослідники отримують диференційовану картину перебігу процесу мовної зміни. Найважливіше, визначається відносна роль контекстуальних особливостей, і сила впливу протягом досліджуваного проміжку часу. Також встановлюється, чи взаємодіяли дві контекстуальні особливості таким чином, що у разі їхньої спільної дії спостерігався відчутний ефект, який був відсутній у разі їхньої ізоляції. Спостереження такого типу дуже важко робити на основі аналізу окремих прикладів, кількісний корпусний аналіз у цьому випадку виконує функцію збільшеного скла, дозволяючи дослідникам виявляти явища, які в іншому випадку могли пройти поза увагою дослідників.

Необхідно зауважити, що висока якість отриманих даних не є самоціллю: точна інформація про процес мовної зміни є важливою передумовою для обговорення причин і характеру мовних змін: зміна може бути результатом дії соціальних чинників (американізація, колонізація), мати культурні або жанрові особливості розвитку (ускладнення, спрощення), бути викликана структурними зрушеннями (узагальнення, аналогічне вирівнювання). Встановлення зв'язків між процесом перебігу зміни і потенційно конкуруючими факторами мовних змін повинно робитися на основі ретельного експертного аналізу всіх чинників. Саме у цьому випадку кількісні методи мають перевагу над якісними: кількісний аналіз дозволяє визначати відносний вплив низки факторів, відокремити важливе від неважливого. Так, аналіз може підтвердити, що певний фактор мав лише незначний вплив і цей вплив можна ігнорувати – щонайменше в аналізі конкретної вибірки та населення, для якого ця вибірка є репрезентативною. Довести це досить переконливо на основі лише якісних даних є досить складним завданням.

Результати, отримані завдяки застосуванню статистичних методів і прийомів, часто вказують на те, що ймовірні моделі мовного використання конструкцій зазнають непомітних на перший погляд змін, які неможливо виявити шляхом порівняння окремих прикладів, проте їхнє значення важко переоцінити в термінах загального процесу мовної зміни. По-перше, варіативний підхід до аналізу використання альтернативних мовних одиниць, що виконують однакову функцію, плідно перенесений на регресивний аналіз варіативності в процесі мовного розвитку. Завдяки діахронічним мовним корпусам, які репрезентують послідовні періоди розвитку англійської мови, час може бути включеним до аналізу як один із взаємодіючих предикторів, також можна порівняти варіативність у сучасності з варіативністю на ранніх етапах мовного розвитку. По-друге, кількісний аналіз пропонує детальний розгляд того, що відбувалося, уможливлючи таким чином визначення конкретних факторів впливу на процес мовної зміни в різні проміжки часу. Динаміка діахронного варіювання може знаходитися у діапазоні від відносної стабільності до суттєвої реорганізації, що потребує високої уваги до подробиць: дослідження повинне не просто дати відповідь на питання про вплив того чи іншого чинника, воно має визначити час початку дії такого чинника і зміни в силі впливу в процесі розгортання мовної зміни. По-третє, результати кількісних досліджень часто вказують на відсутність впливу, що свідчить про принципову невалідність гіпотези, яка передбачала наявність такого впливу.

Потенціал кількісних корпусних методів дослідження в діахронічних студіях полягає також і в експериментальному мікроаналізі (*bottom-up analysis*) [Gries 2011]. Мікроаналіз передбачає використання низки технік, за допомогою яких дані проходять статистичну обробку з метою виявлення мовних одиниць і конструкцій, які не обов'язково передбачалися дослідниками. У цьому випадку на протизагу макроаналізу (*bottom-down analysis*), застосовується зворотній порядок кількісного і якісного аналізу. В той час як логістичний регресивний аналіз потребує підґрунтя у формі якісного аналізу, який пізніше ретельно перевіряється за допомогою статистичних методів, мікроаналіз може починатися зі статистичної обробки первісних даних, результати якої потім використовуються у якісному аналізі. Автоматизовані комп'ютеризовані процедури мають переваги у новизні підходу і можуть використовуватися для подолання упереджень та виявлення тих аспектів досліджуваного явища, які опинилися поза увагою дослідників.

Перегруповання даних для полегшення кількісного аналізу лежить в основі дослідницької аналітичної методики, яка вивчає зміни в синтагматичній поведінці граматичних конструкцій [Gries 2008]. Цей підхід ґрунтується на методі дистинктивного колексального аналізу [Gries, Stefanowitsch 2004], що використовується для протиставлення синтагматичних профілів двох або більше конструкцій з відкритим слотом, в якому можуть актуалізуватися лексичні одиниці різних типів. Ш. Гріс і Ф. Стефанович наводять приклад чергування двох синтаксичних одиниць зі значенням давального відмінку. Дитранзитивна конструкція та прийменникова давальна конструкція часто вживаються з однаковими дієсловами [Gries, Stefanowitsch 2004, с. 106]. Порівнюючи частоту вживання обох синтаксичних одиниць у текстах з частотою дієслів, що вживаються з кожною конструкцією окремо, можна визначити дієслова, що суттєво відхиляються від очікуваної частоти використання. Наприклад, дитранзитивна конструкція частіше актуалізується після дієслів *give, tell* і *show*, в той час як прийменникова конструкція з давальним відмінком переважно вживається після дієслів *bring, play, and take*. Така дистрибуція відповідає припущенню про семантичну різницю між двома конструкціями, яка полягає в дистанції між агентом і реципієнтом (дитранзитивна конструкція передбачає більшу близькість) та тезі про те, що дитранзитивна конструкція в першу чергу виражає, що "X змушує Y отримати Z" в той час як прийменникова конструкція має значення "X змушує Z перейти до Y" [Goldberg 1995, с. 75–76].

Метою колоструктурного аналізу (*collostructional analysis*) є дослідження семантики граматичних конструкцій шляхом вивчення найчастіших синтагматичних "партнерів". При застосуванні цього методу у діахронічних дослідженнях можна визначити і порівняти

синтагматичне оточення однієї і тієї самої конструкції в різних історичних періодах. За допомогою цього методу визначаємо перелік найважливіших синтаксичних і лексичних одиниць, що вживаються з конструкцією в різні періоди її розвитку. Розбіжності в цьому переліку в різні періоди можуть вважатися результатом семантичної зміни: у разі розширення семантичного значення конструкції, вона вживатиметься з більшою кількістю семантично відмінних одиниць; у випадку звуження значення конструкції, вона використовуватиметься з тими ж самими одиницями. В історії англійської мови прикладом першої ситуації є розвиток конструкції *be going to* у період з XVIII по XX ст. [Hilpert 2008, с. 120]. У той час як на початку свого розвитку *be going to* актуалізується з дієсловами, що передбачають використання живих, свідомо діючих агентів в ролі підмета, дані пізніших етапів розвитку цієї конструкції демонструють розширення семантики, вона починає вживатися з дієсловами на зразок *happen*, що використовується на позначення спонтанних подій, а не свідомих вчинків.

Висновки і перспективи подальшого дослідження: кількісні методи мають значний потенціал для діахронних корпусних досліджень, по-перше, тому що для визначення масштабів варіативності, яка є ключовим фактором в процесі мовних змін, необхідно мати аналітичні інструменти, що зможуть забезпечити обробку комплексних даних та дозволять проводити детальний аналіз мовних подій. Для розуміння причин мовних змін, важливо розуміти процес перебігу кожної зміни. По-друге, кількісні аналітичні методи дозволяють приділити увагу тим явищам, які б в іншому випадку пройшли поза увагою дослідників. Ці методи можуть запропонувати свіжий, неупереджений погляд на явища, що здаються зрозумілими. В той же час, необхідно пам'ятати, що навіть високий рівень аналітичного дослідження не допоможе вирішити проблему недосконалості історичних даних, яка є об'єктивним обмеженням в рамках історичних соціолінгвістичних досліджень. У випадку недостатності історичних даних, виходом із ситуації може бути застосування принципу уніформітарності та екстраполяція сучасних моделей варіативності і порівняння їх з історичними моделями.

Література

- Baayen R. H. Analyzing Linguistic Data: A Practical Introduction to Statistics / Baayen, R. Harald. – Cambridge: Cambridge University Press, 2008. – 368 p. Bailey G., Tillery J. Some Sources of Divergent Data in Sociolinguistics / Bailey Guy and Tillery Jan. // Sociolinguistic Variation. – New York: Oxford University Press, 2004. – P. 11-30. Cantos P. The Use of Linguistic Corpora for the Study of Linguistic Variation and Change: Types and Computational Applications / Pascual Cantos // The Handbook of Historical Sociolinguistics. – Wiley-Blackwell, 2012. – P. 99-122. Francis W. N. Language corpora B.C. / Francis W. N. // Directions in corpus linguistics. Proceedings of the Nobel Symposium 82, Stockholm, 4-8 August 1991. – New York: Mouton de Gruyter, 1992. – P. 17-34. Gries S. Th. Commentary: corpus-based methods / Gries, Stefan Th. // Current Methods in Historical Semantics. – Berlin/New York: Mouton de Gruyter, 2011. – P. 184-195. Gries S. Th., Stefanowitsch A. Extending constructional analysis: a corpus-based perspective on “alternations” / Gries, Stefan Th. and Stefanowitsch, Anatol // International Journal of Corpus Linguistics. – 2004. – No. 9 (1). – P. 97-129. Goldberg A. E. Constructions: A Construction Grammar Approach to Argument Structure (Cognitive Theory of Language and Culture Series) / Adele E. Goldberg. – University of Chicago Press, 1995. – 271 p. Hernández-Campoy J. M., Schilling N. The Application of the Quantitative Paradigm to Historical Sociolinguistics: Problems with the Generalizability Principle / Juan M. Hernández-Campoy and Natalie Schilling // The Handbook of Historical Sociolinguistics. – Wiley-Blackwell, 2012. – P. 63-79. Hilpert M., Gries S. Th. Quantitative approaches to diachronic corpus linguistics / Hilpert, Martin and Stefan Th. Gries // The Cambridge Handbook of English Historical Linguistics. – Cambridge: Cambridge University Press. Forthcoming. Hilpert M. Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax / Hilpert, Martin. – Cambridge: Cambridge University Press, 2013. – 248 p. Hilpert M. Germanic Future Constructions: A Usage-based Approach to Language Change / Hilpert, Martin. – Amsterdam/Philadelphia: John Benjamins, 2008. – 208 p. McEnery A., Xiao Z., Tono Y. Corpus-based language studies: An advanced resource book / Anthony McEnery, Richard Xiao, Yukio Tono. – London: Routledge, 2006. – 408 p. Sankoff D., Labov W. On the uses of variable rules / Sankoff David, Labov William // Language in Society. – 1979. – No. 8. – P. 189-222. Tagliamonte S. A. Analysing Sociolinguistic Variation / Tagliamonte, Sali A. – Cambridge: Cambridge University Press, 2006. – 298 p.