

ЛЕКСИКОГРАФІЧНА БАЗА ДАНИХ ЗБІРКИ М. НОМИСА “УКРАЇНСЬКІ ПРИКАЗКИ, ПРИСЛІВ’Я І ТАКЕ ІНШЕ”

© Кульчицький І.М., Осідач Н.Б., 2011

Досліджено структуру збірки М. Номиса “Українські приказки, прислів’я і таке інше” з метою створення її електронної версії. Запропоновано структурну інформаційну модель лексикографічної бази даних та технологію перетворення збірки з паперової форми у цифрову.

Ключові слова: комп’ютерна лексикографія, пареміологія, бази даних.

The article researches the structure of the collection “Ukrainian adages, proverbs and the like” by M. Nomys with the aim of creating its electronic version. An informational model of the lexicographic database and a technology of transforming the collection from hardcopy into a digital version are suggested.

Key words: computer lexicography, paremiology, databases.

Вступ

Закон України “Про основні засади розвитку інформаційного суспільства в Україні на 2007–2015 роки” до однієї з основних стратегічних цілей розвитку інформаційного суспільства в Україні зараховує збереження культурної спадщини України [7, р. I, п. 1]. З цією метою у законі передбачено [7, р. III, пп. 4, 10]:

- створення в електронній формі архівних, бібліотечних, музейних фондів та інших фондів закладів культури;
- формування відповідних інформаційно-бібліотечних та інформаційно-пошукових систем з історії, культури, народної творчості, сучасного мистецтва України тощо;
- переведення в електронну форму національних надбань у сфері культури та мистецтва.

До таких надбань належить збірка Матвія Терентійовича Номиса (Симонова) “Українські приказки, прислів’я і таке інше”. Уперше її видано в Санкт-Петербурзі у 1864 р., у час між Валуєвським циркуляром 1863 року та Емським указом 1876 року, які забороняли українське слово. Заборони вплинули як на саму збірку (цензурування), так і на її розповсюдження вже після видання. Як зазначає М. Пазяк: “...українська збірка прислів’їв Номиса, як і більшість пам’яток української культури, через несприятливі умови життя народу перебувала у напівзабутті і забороні” [10, с. 5].

На початку ХХ ст. М. Возняк знайшов й опублікував у “Записках Наукового товариства ім. Шевченка” (1909, т. 88, кн. 2) [цит. за 10, с. 6] зразки паремій, вилучені цензурою в першому виданні. За радянських часів у 1920-х роках – часі українського відродження – збірку накладом 150 примірників перевидав відомий фольклорист й етнограф А. Лобода. У подальшому М. Номис був зарахований до буржуазних діячів, і збірку в радянській Україні не перевидавали до 90-х років минулого століття, коли журнал “Київ”, починаючи з березня 1991 року, друкує частинами тексти зі збірки Номиса. За кордоном збірку перевидало у 1985 р. з ініціативи та на кошти Владики Мстислава, патріарха Української Автокефальної Православної Церкви, і приурочено до 120-річчя її виходу в світ [10, с. 5–7].

За незалежної України збірку у 1993 р. видало київське видавництво “Либідь”, розпочавши нею серію “Літературні пам’ятки України”, воно ж і перевидало її у 2004 р. у серії “Пам’ятки історичної думки України” [10].

Незважаючи на назву “збірка”, автори роботи, спираючись на енциклопедію української мови (див. статті М. Ф. Алефіренко “Фразеологізм” [9, с. 770], “Фразеологія” [9, с. 774], В. М. Білоноженка “Фразеологічний словник” [9, с. 773], М. Т. Демського “Пареміологія” [9, с. 460], О. О. Тараненка “Словник” [9, с. 610], “Пареміологічний словник” [9, с. 460]) зараховують її до фразеологічних словників та розглядають її як лексикографічну працю, яка має важливе значення для української лінгвістики.

Збірка має низку особливостей, що роблять її унікальною фольклорною пам’яткою. Як зазначають дослідники, М. Номис уперше в українській фразеографії застосував тематичний принцип, поділив розділи на спеціалізовані підрозділи [10, с. 17–18] та запровадив чітку паспортизацію одиниць за місцем їхньої фіксації.

Постановка проблеми

Сучасний стан розвитку лексикографії та застосування у цій галузі новітніх інформаційних технологій дає змогу говорити про реальну можливість інтегровано подати різноаспектну словникову інформацію, тобто досягнути “ідеалу” теоретичної лексикографії. Надбання традиційної лексикографії мають внутрішні ресурси для узагальнення та застосування у нових інформаційних середовищах. Так, наприклад, П. Н. Денісов порушує питання про визнання своєрідним стандартом структур словникових статей, які сформовані у найкращих словниках кожного з типів. Для цього необхідно, по-перше, мати необхідну лінгвістичну інформацію, а по-друге, – розробити формат, зручний для її подання [11, с. 15–16].

Окрім того, один із принципів лексикографічного опису – тяглість традицій, який передбачає, що всякий словник в описі матеріалу завжди спирається на наявні лексикографічні твори [1, с. 7]. Навіть якщо кінцевий варіант словника з різних причин не відповідає первинному задуму укладачів, в подальшому він часто стає корисним джерелом під час укладання інших словників.

З огляду на це, одне із першочергових завдань лексикографії як науки – запровадити в широкий науковий обіг максимальну кількість інформації з репертуару виданих лексикографічних праць. Особливо це важливо для українського словникарства, де велика кількість виданих у кінці XIX – першій половині XX ст. лексикографічних творів стали бібліографічною рідкістю. І хоча більша частина наявної у них інформації втратила свою цінність для широкого загалу, на наше глибоке переконання, вона абсолютно не втратила своєї наукової значущості для науковців різних галузей лінгвістики. У цих словниках законсервоване знання тодішніх учених, що були належними фахівцями у своїх галузях, та тогочасна українська лексика, частина якої сьогодні забута, але вона з успіхом може бути повернута у широкий вжиток [3, с. 3], що має важливе значення для розвитку сучасної наукової термінології.

Повернення лексикографічних праць у науковий обіг можливе двома шляхами: перевиданням друкованим способом та відтворенням в електронній формі. Без сумніву, перевидання повертає такі словники з небуття. Однак з огляду на економічні чинники, їх перевидують не так уже й часто та й невеликим накладом, що не сприяє їхньому широкому розповсюдженню. За приблизним підрахунком, до 1948 р. видано близько 214 словників та проектів словників, до складу яких входить українська мова [2, с. 130–145]. Сьогодні в серії “Із словникової спадщини” їх перевидано 10 (приклад такого перевидання подано в списку літератури [6]).

На нашу думку, продуктивнішим є відтворювати ці лексикографічні і фразеографічні твори в електронній формі. Такий підхід певною мірою дешевший та простіший у тиражуванні. Проте найважливіше – електронна форма урізноманітнює доступ до інформації у цих словниках та підвищує якість її наукового опрацювання.

Вищеподані чинники зумовлюють **актуальність** створення електронної версії збірки М. Номиса “Українські приказки, прислів’я і таке інше”.

Шляхи відтворення словників у електронній формі

За однією з типологічних схем [4, с. 141] електронні словники (словники, записані за допомогою електронних пристроїв на електронних носіях інформації) групують за ознаками “засіб використання” та “спосіб доступу”. За першою ознакою словники поділяють на:

- комп'ютерні – записані та відтворені на персональному або портативному комп'ютері;
- кишенькові – записані на кишенькових електронних пристроях, наприклад, кишенькові перекладачі;

- мобільні – записані в мобільних телефонах.

За другою ознакою виділяють:

- стаціонарні – встановлені на жорсткому диску комп'ютера;
- переносні – записані на компакт-дисках з доступом тільки за їхньої наявності у дисководі;
- мережеві – розміщені на комп'ютері-сервері в мережі й доступні її засобами, наприклад, Інтернет.

Переважно ці словники розраховані на широкого споживача, що зумовлює склад їхньої інформаційної бази, критерії пошуку та структуру відповіді на пошукові запити. Окрім того, не завжди можна встановити джерела інформаційної бази таких словникових систем. Вдалими прикладами таких систем можуть слугувати продукти фірми ABBYY Lingvo, розділ “Словари” сайту російського гуманітарного університету (<http://www.i-u.ru/biblio/dict.aspx>) та проект “Словники України on-line” (<http://lcorp.ulif.org.ua/dictua/>) Українського мовно-інформаційного фонду НАНУ. Компакт-версія останнього містить засоби автоматизованої лематизації словоформ.

Щодо повернення у науковий обіг лексикографічної спадщини, то в цьому напрямі в Україні активно проводять роботу, на жаль, за невеликими винятками, ентузіасти без жодної державної підтримки. Як правило, лексикографічні та фразеографічні раритети сканують та надають доступ до зісканованої інформації у форматах pdf, djvu, doc (див. сайти <http://culture-ua.com/catalog/>, <http://litopys.org.ua/links/inlex.htm>, <http://www.mova.info/>, <http://ukrknyga.at.ua/load/>, <http://mposhuk.com.ua/item/7891236842258/lang/ru>, <http://www.ukrlife.org/main/>, <http://www.madslinger.com/bookvault/index/>), або забезпечують пошук у зісканованій інформації (прикладом може слугувати сайт <http://leksika.com.ua/>). На думку авторів, найплідніше у цьому напрямі працює проект “Російсько-українські словники” [8], де реалізовані обидва способи доступу до інформації. Цікаву збірку паремій подає сайт <http://aphorism.org.ua/>.

У жодному разі не применшуючи необхідність, важливість та корисність наявних електронних форм (автори залюбки користуються ними у своїй повсякденній роботі), зазначимо, що бракує ще однієї форми – текстів словників кінця XIX – початку XX ст. у вигляді інформаційних систем з пошуковим апаратом, який відповідав би сучасним потребам. Це є можливість пошуку слів за сучасною орфографією з одночасним переглядом інформації в орфографії оригіналу, пояснення малозрозумілих сучасному читачеві слів тощо. Ми вважаємо, що таке завдання могли б виконувати кафедри прикладної лінгвістики вищих навчальних закладів України, залучаючи студентів, що опановують цю спеціальність [3].

Саме в такому напрямі на кафедрі прикладної лінгвістики НУ “ЛП” розпочато створення електронної версії збірки М. Номиса “Українські приказки, прислів'я і таке інше”. За основу взято видання збірки видавництвом “Либідь” 1993 року [10].

Мета роботи – проаналізувати структуру збірки, створити її інформаційну модель та запропонувати технологію перетворення збірки з паперової в електронну форму. При цьому автори дотримуються двох основних засад:

- максимальне збереження автентичності збірки;
- максимальне уможливлення наукового опрацювання інформації збірки сьгоднішніми дослідниками.

Аналіз структури та побудова інформаційної моделі збірки

Основою електронної версії лексикографічної праці “Українські приказки, прислів'я і таке інше” буде лексикографічна база даних (ЛБД) [5], куди буде поміщено оригінальну та необхідну додаткову інформацію. Для цього потрібно побудувати інформаційну модель збірки та реалізувати її у концептуальну модель ЛБД.

У структурному аспекті теоретична лексикографія розрізняє макро-, медіо- та мікрорівень словника [12, с. 73]. Розглянемо ці рівні у структурі збірки М. Номиса.

Макрорівень збірки

Макрорівень визначає структуру збірки загалом. За виданням [10] туди входять:

- вступна стаття упорядника збірки М. М. Пазяка;
- передмова М. Номиса;
- приказки та прислів'я – основний корпус;
- варіанти до окремих приказок та прислів'їв основного корпусу;
- загадки;
- покажчик до приказок;
- покажчик до загадок;
- стаття М. Возняка про збірку з вилученими цензурою її складовими;
- примітки упорядника;
- алфавітний посторінковий покажчик опорних слів.

Автентичними складовими збірки є передмова М. Номиса, основний корпус приказок та прислів'їв, варіанти приказок та прислів'їв основного корпусу, загадки та покажчики до приказок і загадок. Розглянемо внутрішню структуру окремих складових.

Передмова М. Номиса під назвою “Од видавця” містить короткий опис історії укладання збірки, способи компресованого запису варіантів приказок та прислів'їв, значення скорочень при вказанні місця фіксації паремії, інші, необхідні, на думку автора, пояснення. На нашу думку, значення скорочень місця фіксації необхідно виділити в окремий об'єкт “Місце фіксації паремії” макрорівня словника, зберігши його текст у передмові для цілісності останньої. Цей об'єкт буде використано як довідник при описанні паремій.

Автентичні покажчики до приказок і загадок необхідно, на нашу думку, відобразити без змін у ЛБД, незважаючи на те, що їх можна автоматично отримати з корпусу паремій. Пізніше їх можна буде порівняти з автоматично сформованими.

Стаття М. Возняка містить короткий коментар автора та паремії, що були вилучені зі збірки М. Номиса цензурою. На нашу думку, ці паремії доцільно долучити до основного корпусу паремій збірки з вказанням джерела.

Вступну статтю М. Пазяка, статтю М. Возняка на макрорівні доцільно об'єднати у єдиний об'єкт “Про збірку”, до якого у подальшому можна буде долучити інші статті подібного змісту.

Про примітки упорядника буде сказано нижче, а алфавітний посторінковий покажчик опорних слів у ЛБД зайвий, позаяк має відношення лише до паперового перевидання збірки.

Отже, на макрорівні збірка містить такі об'єкти:

- передмову М. Номиса;
- корпус (пронумерований перелік) паремій, який, своєю чергою, містить:
 - паремії з автентичної збірки Номиса (назвемо їх основним корпусом);
 - паремії, що опубліковані у статті М. Возняка (назвемо їх додатковим корпусом);
- покажчики, які міститимуть:
 - авторський покажчик до приказок та прислів'їв;
 - авторський покажчик до загадок;
 - покажчик до прислів'їв та приказок, побудований автоматично;
 - покажчик до загадок, побудований автоматично;
- довідник, який містить відповідності між скороченою та повною назвою місця фіксації паремії;
- об'єкт “Про збірку”, у якому буде поміщено:
 - передмову М. Пазяка;
 - статтю М. Возняка тощо.

Макрорівень структури збірки показано на рис. 1.

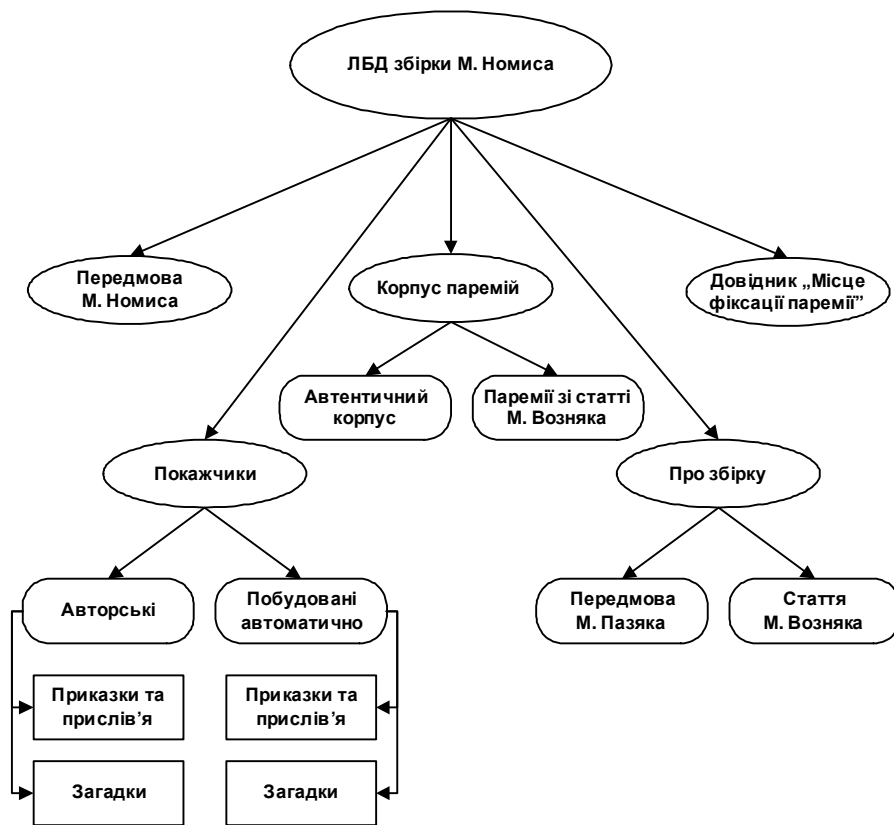


Рис. 1. Макроструктура збірки М. Номиса.
Медіорівень збірки

Медіорівень словника відображає взаємозв'язки між окремими статтями словника. Розглянемо їх у збірці М. Номиса. Насамперед (див. опис макрорівня), усі статті збірки будуть зараховані до двох груп: збірка М. Номиса та стаття М. Возняка.

Основний корпус одиниць автор збірки згрупував у 239 тематичних груп, кожна з яких названо та подано номери паремій, що належать до цієї групи (наприклад, *Багатир* – 1 343–1 463 [10, 100]; *Природу (прирожденне) тяжко одмінить* – 3 208–3 232 [10, 141]; *Битий жак* – 5 704–5 774 [10, 269]; *Смерть* – 8 240–8 338 [10, 365] тощо). Окрім тематичних груп, подано дві структурні: “Приказки и д. невідомі видавцеві” [10, 576] та “Додаток” [10, 581]. Тематичні групи розбито на розділи, назв яким автор не давав, а просто уміщав список відповідних груп на початку розділу (наприклад, *Кохання* – 8 726–8 785; *Роспушта* – 8 786–8 844; *Женитися* – 8 845–9 018; *Чоловік і жінка* – 9 019–9 166; *Діти* – 9 167–9 283; *Син* – 9 284–9 310; *Дочка* – 9 311–9 330; *Рід* – 9 331–9 417 [10, 365]).

Отже, кожна паремія зарахована до тематичної або структурної групи, кожна з яких належить до відповідного розділу. У лексикографічній базі даних розділам будуть просто присвоєні номери. Медіоструктуру збірки показано на рис. 2. Під основним і додатковим корпусом паремій слід розуміти збірку М. Номиса і статтю М. Возняка.

На мікроструктурному рівні розглядають структуру словникової статті. У збірці фразеологічні одиниці подано пронумерованими гніздами. Узагальнена структура гнізда така:

- номер (в оригіналі, мабуть, з метою економії, зазначено номери з певним кроком – або кожне десяте (гніздо), або там, де гніздо вилучив цензор);
- фразеологічна одиниця з варіантами;
- коментарі та пояснення;
- місце фіксації паремії чи її варіантів.

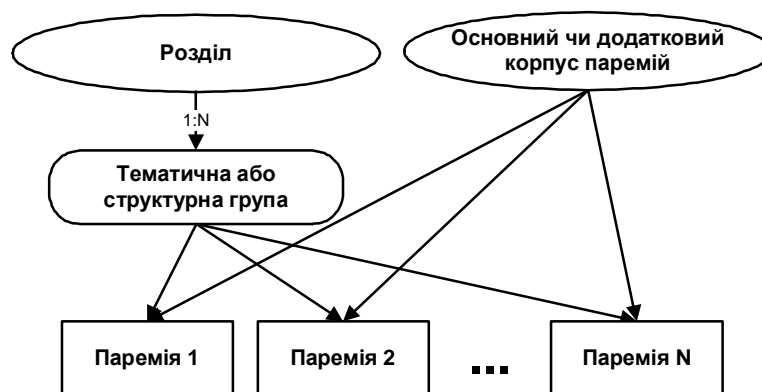


Рис. 2. Медіоструктура збірки
Мікрорівень збірки

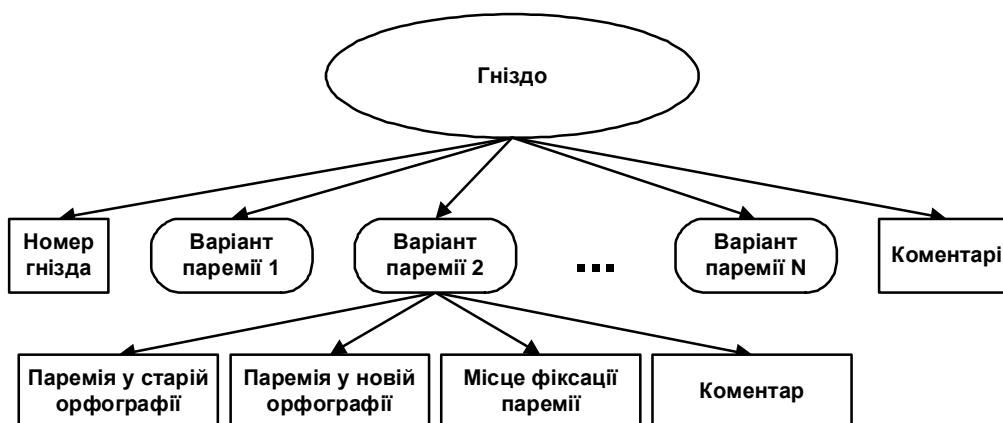


Рис. 3. Мікроструктура збірки

Паремії у словнику написані у старій орфографії. На нашу думку, подання словникових гасел лише в оригінальному варіанті утруднить, а подекуди й унеможливить сучасному користувачеві автоматичний пошук у словнику за компонентами фразеологічних одиниць. Тому у базі даних фразеологічні одиниці будуть продубльовані сучасною орфографією. Мікроструктуру збірки показано на рис. 3.

Технологія перетворення збірки в електронну форму

Для перетворення збірки М. Номиса з паперової у електронну форму необхідно виконати такі роботи:

- перетворення тексту збірки в електронну форму, придатну для опрацювання її текстовими редакторами;
- звіряння електронної форми збірки з оригінальним текстом та виправлення помилок;
- структуризація інформації. З цією метою на ґрунті інформаційної моделі буде розроблена система маркерів, які будуть внесені у текст збірки;
- розроблення та реалізація концептуальної моделі лексикографічної бази даних. Буде використано промислову реляційну систему керування базами даних;
- розроблення спеціального програмного забезпечення для автоматичного переносу автентичного маркованого тексту збірки у базу даних;
- розгортання скорочених форм запису інформації.

Скорочені форми запису інформації автор, мабуть, застосовував з метою економії місця. При перенесенні інформації у базу даних ці скорочені форми необхідно розгорнути. При цьому труднощі виникають для місця фіксації та варіантів паремій. Розглянемо ці проблеми докладніше.

Місце фіксації паремії

У передмові укладач подає скорочення та повні відповідники чи назви місця запису фразеологізму чи імені того, хто подав одиницю [10, с. 31–35]. Цю інформацію належно буде перенесено у базу даних. Скорочення будуть інтерпретовані у такий спосіб (за вказівками автора та власною логікою):

“Де при приказці (напр. № № 2855, 4826), або при одміні (одміна під ціхвірею 2 в № 4824) нема ніякого напису, або поставлено id, там треба розуміть напис, що и при передні приказці, або одміні...”:

У найпростіших варіантах все зрозуміло.

Приклад 1. Написано у збірці [10, 236]:

4825. *И сам не ість, и другому не дасть. Ил.*

4826. *А ні нам, а ні вам.*

Паремії 4826 приписуємо “Ил.”

Приклад 2. У збірці написано [10, 236]:

4824. *Собака на сїні лежить* ⁽¹⁾,— *и сама* ⁽²⁾ *не ість, и другому* ⁽³⁾ *не дає* ⁽⁴⁾. Л., Ёвх., [Проск.].— *Лежить собака на стерві,— и сам не буде істи, и другому не дасть.* Бр., Бер.

⁽¹⁾ *Лежачи. Ст. Зб.* ⁽²⁾ *ні сам. id; и сам. Пр., Тр.* ⁽³⁾ *а скотові. Пр.* ⁽⁴⁾ *ні товарові не дає істи. Ст. Зб.*

Варіантові ⁽²⁾ приписуємо “Ст. Зб.”

Приклад 3. Наступний приклад викликає неоднозначне трактування. Написано у збірці [10, 160]:

2854. *ЛИХИЙ* ⁽¹⁾ *доброго не любить.* Ст. Зб.

⁽¹⁾ *Лихе. Ил.; Зле. Бр.*

2855. *Як лихий чоловік, то не став и чорта.*

Неоднозначність у тому, що у цьому випадку розуміти під “передньою приказкою або одміною”. Якщо виходити з того, що після розкриття компресії отримаємо таку послідовність: *лихий доброго не любить Ст. Зб., лихе доброго не любить Ил., зле доброго не любить Бр.*, то до паремії 2855 має бути приписано “Бр.”. Якщо ж варіанти паремії 2854 вважати чимось внутрішнім та проаналізувати лексичний склад паремій “*лихий доброго не любить*” і “*як лихий чоловік, то не став и чорта*”, то до 2855 має бути приписано “Ст. Зб.”. Проблемою є те, що обидва варіанти мають право на існування. Тому такі випадки при внесенні інформації в базу даних вимагатимуть додаткового аналізу та відповідної примітки про неоднозначність.

Приклад 4. “...Де напис стоїть в ламаних скабочках (напр. при № № 4750, 4801 [Бр.], [Гр.]), треба уважать, що до приказок мались и одміни (з Бр. и Гр. в вименених нумерах), але ті одміни нічим не примітні, ніякої користи не вбачалось друкувать їх...”: **4750.** *Дав би на молебень, та й собі потребен.* Рад., [Бр.] [10, 234]; **4801.** *Що тому богу молицьця, которий не милує.* Пир., [Гр.] [10, 236].

Приклад 5. “...Де в написах перед менням повіту и т. и. стоять римські ціхвірі, то для того, щоб одрізнить кільки одмін з одного місця (напр. при № 2597 є Бр. и I Бр.— щоб одрізнить одну брацлавську одміну од другої брацлавської) [10, 148]...”:

2597. *Не вважайте* ⁽¹⁾, *люде добрі, що я швецъ: говоріть зо мною* ⁽²⁾ *як з простим* ⁽³⁾. Бр.— *Чуєш, говори зо мною, як з простим, а не думай того, що я швецъ. Не.*

⁽¹⁾ *Не вжахайтєся. I. Бр.; Не потурай. Бр., Проск.; Не думай. Коз.* ⁽²⁾ *говоріть. I. Бр.; говори. Бр., Проск., Коз.* ⁽³⁾ *з простим чоловіком. I. Бр., Проск.; до простого. Бр.*

Ці два випадки при перенесенні в базу даних труднощів не викликають. Їх буде зафіксовано у базі даних за допомогою спеціальних приміток.

Розгортання компресованих форм подання паремій у збірці

Пояснюючи спосіб компресії варіантів паремій, автор у передмові подає такий коментар [10, с. 35]: “...Де перед одміною маецця кільки цяток (...), то, чи будуть ті одміни надруковані під приказкою дріботою, чи надруковані ик строці з приказкою, треба вважать їх яко кінєць переднєї

одміни чи приказки, та ще й так: коли в одміні маєця яке слово, що маєця и в передні одміні чи приказці (напр. № 2815, ик строці, маєця в одміні слово жінка), то виходить, одміну сю так треба читать: “Коли моя жінка шельма и д.”; де такого слова нема (№ 2782, 3-я одміна) – так: “Приший-кобилі-хвіст, а в кобилі и свій є”, або (там же, 5-а одміна): “Приший- хвіст-кобилі: як одирвеця, то далеко опинеця”; де, в одміні, стоїть значок (“),— треба вважать, що в одміні сі нема слова, до якого значок поставлений (напр. під № 2823 в Лубенські и Илькевичевські одміні нема слова блювати).”

Під час розгортання компресованих форм одиниць та “прочитання” варіантів виникають такі труднощі:

коли варіант позначено цифрою, то вона не завжди стосується лише одного слова зліва, біля якого написана;

коли одиниця має кілька варіантів під різними цифрами, не завжди очевидним є лексичний склад конкретного варіанта.

Для утворення правильного списку варіантів одиниць, на нашу думку, необхідно уважно проаналізувати текст та стежити за вказанням місця запису варіанта одиниці у скороченому варіанті. При цьому необхідно враховувати те, що в тексті оригіналу можливі неточності. Розгляньмо цей процес на прикладах.

Приклад 6. У збірці подано [10, 158]:

2818. *He* ⁽¹⁾ *по коню* ⁽²⁾, *та* ⁽³⁾ *по голоблях* ⁽⁴⁾. *Лів.*

⁽¹⁾ *He може. П., Прав. Ниж., Кан., К.; He міг. Проск.; He можь. Ил.* ⁽²⁾ *по коняці. П., Прав. Ниж.; по коневі. Проск.; He понавсь мене. Бр.* ⁽³⁾ *та понав. Бр.; то. Ил.* ⁽⁴⁾ *голоблях. Бр.; Проск., оглоблях. Пол.; голоблі. Сл.; кульбаці. Ил.*

Для утворення розгорнутого списку варіантів цієї паремії утворюємо з основної форми та її варіантів таку таблицю:

1	2	3	4	Місце
He	по коню	та	по голоблях	Лів.
	He понавсь мене	та понав	по голоблях	Бр.
He можь		то	по кульбаці	Ил.
He може				К.
He може				Кан.
He може	по коняці			П.
			по оглоблях	Пол.
He може	по коняці			Прав. Ниж.
He міг	по коневі		по голоблях	Проск.
			по голоблі	Сл.

Відразу можна зауважити, що в оригіналі для місцевості “Бр.” вкралася описка – мабуть, достатньо було вказати “понавсь мене”. Якщо в таблиці порожні місця заповнити відповідними словами з основної паремії, то отримаємо такий перелік паремій гнізда 2818:

He по коню, та по голоблях. Лів.

He понавсь мене, та понав по голоблях. Бр.

He можь по коню, то по кульбаці. Ил.

He може по коню, та по голоблях. К.; Кан.

He може по коняці, та по голоблях. П.

He по коню, та по оглоблях. Пол.

He може по коняці, та по голоблях. Прав. Ниж.

He міг по коневі, та по голоблях. Проск.

He по коню, та по голоблі. Сл.

Приклад 6. У збірці подано [10, с. 41]:

58. *Бог старий господарь*. Бр. – ...має більше, ніж роздасть. – *Більше Бог має, як роздав*.
Ил., Кан., К.

Отримуємо:

Бог старий господарь. Бр.

Бог старий господарь має більше, ніж роздасть. Бр. (за правилом невказаної місцевості)

Більше Бог має, як роздав. Ил., Кан., К.

Приклад 7. У збірці подано [10, 158]:

Коли моя жінка така, то нехай свині борошно ідять. Рад., Пир.—...жінка шельма, нехай
(¹) *свині муку ідять*. Б.

(¹) *к...а, так нехай же*. Гр.

Отримуємо:

Коли моя жінка така, то нехай свині борошно ідять. Рад., Пир.

Коли моя жінка шельма, нехай свині муку ідять. Б.

Коли моя жінка к...а, так нехай же. Гр.

Приклад 8. У збірці подано [10, 156]:

Приший хвіст! Гр.— *Приший-кобилі-хвіст!* Коз., Кр., [Бр., Проск.]—... *а в кобили и свій є*.
К., Рад.— *Приший-хвіст-кобилі* (¹). Лів.—...*як обирвеця, то далеко опинеця*. Л.— *Причепить*
кобилі білий хвіст. Кулж.— *Причепи кобилі хвіст*. Ил.

(¹) *Кобилій*. Ст. Зб.

Після розгортання отримуємо:

Приший хвіст! Гр.

Приший-кобилі-хвіст! Коз., Кр., [Бр., Проск.].

Приший-кобилі-хвіст, а в кобили и свій є. К., Рад.

Приший-хвіст-кобилі. Лів.

Приший-хвіст-кобилій. Ст. Зб.

Приший-хвіст-кобилі, як обирвеця, то далеко опинеця. Л.

Причепить кобилі білий хвіст. Кулж.

Причепи кобилі хвіст. Ил.

Приклад 9. У збірці написано [10, 159]:

2823. *Нічим* (¹) *вовкові* (²) *блювати* (³), *так* (⁴) *ликами* (⁵) *Пр*.

(¹) *Немає чим*. Ил., Проск. (²) *вовк*. Ил., Пр. в Ст. Зб. (³) *смердіти*. Проск.; *с...ь*. Л., Новг.,
вовка рвать. Гл.; Ил., Л. (⁴) *то* Ил.; *та*. Проск. (⁵) *ликом, id.; завертами (завертнями)*. Новг.

На жаль, під час передруку оригіналу (видавництво “Либідь”) значок “””, яким упорядник позначав відсутність слова у варіантах, втрачено, тому цю ознаку можна визначити за “;” в переліку місцевостей. Залишається сподіватися, що крапка з комою завжди стоїть правильно. Окрім того, під час аналізу скороченого запису варіантів фразеологізмів виникає колізія з місцевістю, позначеною як “Л.”, де у варіанті (3) подано дві видозміни “с...ь” і відсутність слова “блювати”. Знаючи скрупульозність укладача, залишається припустити, що в оригіналі механічно пропущена римська цифра “I” для позначення двох варіантів з однієї місцевості. Якщо так, то отримаємо такий перелік паремій:

Нічим вовкові блювати, так ликами. Пр

Нічим вовкові вовка рвать, так ликами. Гл.

Немає чим вовк, то ликами. Ил.

Нічим вовкові, так ликами. І. Л.

Нічим вовкові с...ь, так ликами. Л.

Нічим вовкові с...ь, так завертами (завертнями). Новг.

Нічим вовк блювати, так ликами. Пр. в Ст. Зб.

Немає чим вовкові смердіти, та ликом. Проск.

- Видрук інформації з бази даних у формі оригіналу збірки.

- Звіряння отриманого тексту з оригіналом.
- Усунення виявлених неточностей;
- Доповнення бази даних необхідною інформацією. До такої інформації зараховано: запис паремій у сучасній орфографії та необхідні коментарі і тлумачення.

Коментарі та тлумачення

Коментарі у збірці подані доволі скупо, це або пояснення значень окремих слів (*174. Те іде молицьця, а те живицьця* (крати). Зал. [10, с. 45]), або етимологічні зауваги (*1063. Як мисль, так мисль – таки буде Перемишль*. Ил. Стародавній галицький город над Саном. Славлять, що слова сі сказав цареві простий Русин, як той добірав мення городові [10, с. 88], або тлумачення значення фразеологічної одиниці (*176. Святий та Божий! свічки поїв, а поночі сидить*. Кон., К., Евх. Як не по правді робить, та ще й виправляєця [10, с. 45]). На нашу думку, оскільки до сьогодні в українській фразеології немає етимологічного фразеологічного словника, то належить, з одного боку, зберегти оригінальні коментарі, а з іншого, – не порушуючи структури збірки, значно доповнити коментарі сучасною інформацією. Зокрема, уточнення та пояснення потребують такі компоненти:

- антропоніми (*Мстислав не одного стиснув* [10, с. 72]; *Старший Гриць, як Парася* [10, 86]; *Пан з паном, а Иван з Иваном* [10, с. 87];
- міфоніми (*Упир і непевний усім відьмам родич кривний* [10, с. 47]; *За царя Гороха, як було людей троха* [10, с. 316]);
- біблійні оніми (*Чим той Лазар Богу приподобився, що ёго увесь тиждень тнуть та й тнуть!* [10, 45]; *Трусиця, як Каїн* [10, с. 173]);
- зооніми (*Пожививсь, як сірко з паски* [10, с. 118];
- гідроніми (*Як ся зійде Стрий і Ломець (ріки), то буде світу кінець* [10, с. 56]; *Дон, Дон, а дома луче* [10, с. 74]);
- астіоніми (*Золотоноша кругом хороша* [10, с. 72]; *В Станіславі кождий на свої справі* [10, с. 72]);
- флороніми (*Коли б не тирлич, був би я твій панич* [10, с. 47]; *Щоб тебе понесло на папороть-могили та на Щокотин слід* [10, с. 592]);
- фауноніми (*Чумак чумака таранею допіка, а сам у ёго з воза потягує чабака* [10, с. 590]; *Пан лупить хлопа, як скопа (валаха), а дідько пана, як барана* [10, с. 697]);
- артефакти (*Піп живи з олтаря, а писар з каламаря* [10, с. 47]; *Доц лле, як з луба* [10, 65]; *Як маєш кланяцьця лаптю, то лучче поклинись чоботу* [10, с. 87]);
- застарілі одиниці виміру (*Коли маєш сто кіп, то будеш піп* [10, с. 46]; *Правди на ніготь, а прибавиця на лікоть* [10, с. 322]);
- релігійні свята, пости тощо (*З Сорока Святих сорок морозів буде* [10, с. 57]; *Петрівка на хліб катівка* [10, с. 60]).

Необхідно також подати інформацію про стереотипні уявлення тогочасної української лінгвостільності про представників інших етносів, оскільки у збірці вміщено значну кількість одиниць із зазначеним компонентом (*Невірний гіри Жида, або Турка* [10, с. 39]; *Постимо, як Рахмани* [10, с. 43]; *Від москаля поли вріж та тікай* [10, с. 76]; *Лях тоді добрий, як спить; а пробудиця, то біда* [10, с. 78]; *Московський час, як жидівський зараз* [10, с. 78]; *Грошей, як у Жида* [10, с. 100]).

Крім коментарів до окремих компонентів фразеологічних одиниць, належить за необхідності умістити також тлумачення одиниць, оскільки значення частини фразеологізмів важко розшифровують сучасні мовці. Наприклад, одиниця *Набожний – як Жид подорожній* [10, с. 45] має значення “зовсім не набожний”, однак в оригінальному тексті не подано тлумачення. Аналогічно: *У мене стільки грошей, як у Жида свиней* [10, с. 105]; *Горобина ніч* [10, с. 65] та ін.

При внесенні коментарів може бути використана інформація з коментарів упорядника паперового видання збірки М. Пазяка.

Висновки

Побудовано структурну інформаційну модель збірки М. Номиса “Українські приказки, прислів’я і таке інше”. Запропоновано технологію перетворення збірки з паперової у електронну форму. Запропонована технологія може бути використана для створення електронних версій видань подібного типу.

1. Дубічинський В. В. *Лексикографія української мови: Конспект лекцій* / В. В. Дубічинський. – Харків: НТУ “ХПІ”, 2002. – 55 с. 2. Кубайчук В. П. *Хронологія мовних подій в Україні: зовнішня історія української мови* / Віктор Павлович Кубайчук. – К.: “К. І. С.”, 2004. – 176 с. 3. Кульчицький І. М. *Бази даних як засіб повернення у науковий обіг української лексикографічної спадщини* / І. М. Кульчицький, І. О. Ліхнякевич, Е. А. Калниня // *Українська термінологія і сучасність: зб. наук. праць / відп. ред. Л. О. Симоненко.* – К.: Інститут української мови НАН України, 2009. – Вип. 8. – С. 273–275. 4. Кульчицький І. М. *Деякі аспекти типології словників* / І. Кульчицький, А. Костенко, Н. Осідач // *Прикладна лінгвістика та лінгвістичні технології: MegaLing-2008: зб. наук. пр. / НАН України, Укр. мовно-інформ. фонд [та ін.]; редкол.: Ю. Д. Апресян [та ін.].* – К.: Довіра, 2009. – С. 135–145. 5. Кульчицький І. М. *Комп’ютерно-технологічні аспекти створення сучасних лексикографічних систем* / І. М. Кульчицький. – К.: НБУ ім. В. І. Вернадського, 2002. – 59 с. 6. *Практичний російсько-український словник приказок: відтворення видання 1929 року* / Інститут енциклопедичних досліджень НАН України; упоряд. Г. Млодзинський; за ред. М. Йогансена. – К.: Ін-т енциклопедичних досліджень НАН України, 2009. – 108 с. – (із словникової спадщини; Вип. 8) 7. *Про основні засади розвитку інформаційного суспільства в Україні на 2007–2015 роки: закон України від 9 січня 2007 року № 537-V* // *Відомості Верховної Ради України.* – 2007. – № 12. – С. 102. 8. *Російсько-українські словники [Електронний ресурс].* – Режим доступу: <http://www.r2u.org.ua/>. 9. *Українська мова: енциклопедія* / редкол.: В. М. Русанівський (співголова), О. О. Тараненко (співголова), М. П. Зяблюк та ін. – 2-ге вид., випр. і доп. – К.: Вид-во “Укр. енцикл.” ім. М. П. Бажана, 2004. – 824 с. 10. *Українські приказки, прислів’я і таке інше: Уклав М. Номис / упоряд., приміт. та вступна ст. М. М. Пазяка.* – К.: Либідь, 1993. – 768 с. (“Літературні пам’ятки України”). – К.: Либідь, 2004. – 352 с. (“Пам’ятки історичної думки України”). 11. Широков В. А. *Елементи лексикографії* / В. А. Широков. – К.: Довіра, 2005. – 304 с. 12. Широков В. А. *Феноменологія лексикографічних систем* / В. А. Широков. – К.: Наук. думка, 2004. – 328 с.