

УДК 930.25:784.4:398(477)(71)

**BUILDING UKRAINIAN FOLKLORE DATABASES:  
COMMUNICATING INFORMATION  
TO AN INTERNATIONAL AUDIENCE**

**Natalie KONONENKO**

*Peter & Doris Kule Centre for Ukrainian & Canadian Folklore,  
Modern Languages and Cultural Studies, Faculty of Arts, University of Alberta  
441-C Old Arts & Convocation Hall,  
tel.: 780-492-6810, e-mail: natalie.kononenko@ualberta.ca*

Folklore fieldwork generates large quantities of data. The scholar's task is to process this data to suit the maximum number of users. Work on websites featuring Ukrainian folklore began in 2000 and led to a better understanding of the differences between researcher needs and what students and the general public require to use a site effectively. Web presentation offers special advantages when it comes to audio because sound captures the nuances of speech. The problem is building databases that present folklore audio in a way that satisfies both researcher and public needs. The Ukrainian Folklore Audio website was among the first to take the user directly to the point in a recording where a chosen topic is discussed. Current coding practice uses International Ethnographic Thesaurus metadata and the Sanctuary Project sound file database of Ukrainian Canadian recordings follows this system. Sanctuary Project encoding work has led to the expansion of the Ethnographic Thesaurus to include more Ukrainian subject matter.

*Key words:* folklore fieldwork, digital technologies, Ukrainian folklore in Canada, The Ukrainian Folklore Sound Recordings website, International Ethnographic Thesaurus.

I began doing ethnographic fieldwork in Ukraine in 1998. As a trained folklorist, I had wanted to do fieldwork all along, but could not. Ukraine's status as a part of the Soviet Union restricted travel and controlled access to data. It was only when the Soviet Union fell apart and Ukraine became an independent country that real fieldwork, especially on the part of foreigners like myself, could begin. I had worked in Ukraine earlier in my career. During Soviet times I had worked in archives, producing a book on Ukrainian minstrelsy, an artistic tradition that had virtually disappeared by the time I collected my data [2]. Studying traditions from the past was the only option during that period because working with contemporary phenomena was anywhere from difficult to impossible under Soviet rule. With the independence of Ukraine, I was able to travel to rural areas and to collect information about village life. I was able to interview people about rituals such as weddings and baptisms and to observe ritual practices. The excitement of documenting material that had been hidden for so long made me eager to share it with my students. This, plus my interest in digital technologies, led to my desire to create a Ukrainian folklore website.

At the time of this work I was at the University of Virginia, a leader in humanities computing. In 2001, with the help of University funding and with University technological support, I created **The Ukrainian Village Project**. This consisted of a database of

photos to be used for research purposes and a website for use in my classes and to deliver information to the general public. My original plan was build a database and that alone. I thought that a database was all that I would need to make my work widely accessible. Building a database taught me a great deal. My original assumption was that online delivery of information would allow me to communicate the complexity of folklore more fully than I could through writing an article or a monograph. The opposite turned out to be true. The relationship between areas of folklore is multifaceted. Material objects such as *rushnyky* are used to decorate and protect the home and they are also ritual objects, used in weddings and funerals. Lyric songs express emotion. Historical songs express emotion as well, but they also record and describe important events. When I built my first database, one that consisted of photographs and descriptions, I tried to capture the complexity of folklore – and failed in communicating the nature of Ukrainian culture and lore to my students. The database was useful to me because I was already familiar with Ukrainian folklore and its interrelationships. When I used my databases in class, however, I discovered that it was far too complex for people with no prior knowledge to use effectively. Presenting materials to students with no background and no knowledge of Ukrainian culture, I learned, had to be done in a simplified way. The complex relationships that I was able to capture in digital form were beyond the comprehension of students starting out in the field. To teach my students effectively, I needed to start with something simple and only then build toward representations of greater complexity. What I ended up doing was keeping my database for my own research and developing a website with selected images and simple text for my students, a website that presented limited information about Ukrainian folklore. This website became known as the **Ukrainian Village Project** and was indeed a useful teaching tool. It was also a site that got good responses from the general public. That site has been updated on an ongoing basis. It is now called **Folklore Ukraine** and can be seen at <http://www.artsrn.ualberta.ca/folkloreukraine/>.

Major updates to the Folklore Ukraine site came at the University of Alberta. I joined the faculty of this institution in 2004 and, with the encouragement of the Arts Resource Centre technical staff and with funding provided by the Centre for Teaching and Learning, I worked on the folkloreukraine website and built other websites. Creating public websites is a service to the community and a wonderful way to present one's work. But I wanted more. I still thought of computer assisted data management and retrieval. I was fortunate to have Svitlana Kukharenko as my graduate student. Kukharenko and I began indexing the many sound files that I had recorded in Ukraine. My intention was to make all of my sound recordings accessible in searchable form. The process was complex. Most of the recordings were on cassette tapes and these needed to be digitized. The value of digitization is obvious. Because cassettes are physical, rather than virtual, objects, finding a particular passage in a cassette recording means winding the tape back and forth until the desired point is found. Moving a cursor through a digital object requires no manipulation of a physical piece of tape and increases searching speed enormously. Still, the process of finding material, even in a digitized recording, can be laborious. By the time that I started trying to index my audio files and to make them searchable, I had collected over one hundred hours of sound. Because my recordings were originally on cassette, they were divided into units that were

forty-five minutes long and there were many of these. Searching through some two hundred files was time-consuming. My careful annotation of my data was of some help but, considering the volume of material in my possession, the annotation proved inadequate for rapid retrieval. In my annotation I had noted the date, place, and time of recording. I had noted the people who were interviewed and the scholars present during the interview session. The content on the recording was also listed in my records. That is, I indicated those tapes where the respondent had talked about courtship practices, or marriage rites, or funeral ceremonies. I had noted the narration of legends or tales. All of my notes were typed, digitized, and thus easily searchable. Still, all they did was to take me to the appropriate forty-five minute sound segment. They did not show me where, in the particular forty-five minutes of data, the material that I needed could be found.

One way to make the searching of field data easier and faster is to transcribe all of the sound files. In Ukraine, the processing of sound recordings was done precisely this way. Transcribing means that a person listens to the recording and writes down everything that is said. Having a transcription of a recording is very useful. It creates a text file and a text file, especially one that is in digital form, can be easily searched using a commonly available search engine such as the one provided with MSWord and other word-processing software. Having searchable text greatly reduces the time needed to find a particular passage. There are two drawbacks to transcription, however. One is financial. In the West in particular, finding and paying for the human resources needed to complete one hundred hours of transcription was out of the question. The other drawback is loss of the actual sound and all of the aural clues to meaning which it provides. Folklore is oral. Folklore is also art. Much of what folklore communicates is expressed by artistic means. Narratives, for example, are characterized by voice modulation: shifts in tone, volume, speed, and so forth. Rendering these in print is difficult. Some attempts have been made in this regard as can be seen in the book, *Prozovyi fol'klor sela Ploske* published by my colleague Oleksandra Britsyna [3]. Here slower tempo is shown by an increase in spacing between letters; louder volume is indicated by boldface type; emphasis is marked by italics, and so forth. Such elaborate transcription, while useful, is even more time-consuming to do than regular transcription. It also increases publication costs, a problem that has grown increasingly acute in the West. Furthermore, it does not fully capture the expressive potential of actually hearing sound. There are some full transcriptions of a few large group projects that have been done in the West. One example is the British National Corpus (BNC). It is a 100 million word, searchable, collection of samples of written and spoken current British English. Yet even this massive work has its problems and many sections are not transcribed, but marked as "unclear".

The obvious desired form of preparing field recordings for research is to work directly with sound and to provide annotation on the sound files themselves. What my technical support staff, Kukharenko, and I envisioned doing was to index the sound files that I had produced and to use that index as metadata, allowing the researcher to go directly to the point in the actual sound recording where a particular type of information was being discussed.

Indexing means noting all of the topics discussed in a recording and the exact time point where the discussion occurs. Indexing requires a template, a “Master Interview Index Form”. This form has to be tailored to the types of information in the interviews. My interests are in wedding, funeral, and birth customs; therefore, these were designated as the major categories. Sub-categories, at that point in our work, were the steps of the rituals designated as the major categories. The process of indexing requires listening to a sound recording and noting the time when one of these categories or sub-categories (keywords) is mentioned. The noting of the keyword and the time point is done on the “Interview Index Form” for that interview. The top part of such a form lists the name of the recording and the date, place and people being interviewed. Once the template has been made, and the template can be modified as one works, indexing is relatively quick. One does not need to go back and forth and listen several times the way one does with transcription. Indexing can generally be done on a single, almost real-time, pass through the material, especially once the indexer is familiar with the template. For each sound file, indexing produces a list of the template categories that are present in that recording, plus the times in the recording when the topics are discussed. This was done with the original set of 150 hours of mp3 recordings and the total Index was a 152 page Word document. For example, three consecutive entries indexed in the recording ploske2000s.mp3 are:

04.04–06.25	весілля – чесна/нечесна молода;
06.26–06.48; 12.13–12.58	весілля – сніданок для молодих;
14.56–16.26	весілля – ділять коровай/дарування.

This document was very useful to me as it could be searched for any specific topic. The search would show in which mp3 files and at what times that topic was discussed. The mp3 files could then be listened to from that time point. As my work progressed, I wanted to make the search process more “automatic”, I wanted to have a “link” between the various topics and the appropriate section of the corresponding sound file. From the above example, three sections need to be linked for such an automatic search to be possible. The three sections are: the name of the recording (1), the times (2), and the topic (3).

- 1) ploske2000s.mp3;
- 2) 06.26–06.48; 12.13–12.58;
- 3) весілля – сніданок для молодих.

We initially intended to cut the original mp3 into small segments, one for each topic shown (for example: “ploske2000s 06.26–06.48”) and to link each segment to the topic. That way, when a topic was searched for and found, the program would play the corresponding small segment. This approach has three problems: first, approximately 5,000 segments would have had to have been cut out; second, if there were a small error in the index times, the segment would then be useless; and third, the context of the segment in the whole recording would be lost. Thankfully, we did not have to cut because programmer Eric (Yue) Zhang, who worked in the Text Analysis Portal for Research at the University of Alberta, wrote a search program and showed us how to “link” topics to the sound segments without cutting. The data which the program searches is in an XML file which we generated from the data in the original “Master Interview Index Form” Word documents.

Unfortunately, when the XML file we had generated was uploaded to the server where the search program resided, the program did not work. There was an error in the XML code. Although we have described this process in a few words, the process of converting the data in the 152 page Word document into the 23,129 line XML file was not simple. We used several computer programs we wrote as well as the many “find-and-replace” functions of Word to convert it into the XML format. Mozilla Firefox, which shows formatting errors in XML files when it attempts to open them, unfortunately, did not see any obvious formatting errors. The explanation was the visual similarities between certain Latin and Cyrillic letters and the differences in their computer encoding. Although the letters: **a**, **c**, **o** etc. look identical in English and Cyrillic, the former are encoded in the computer as ASCII (one byte per letter) whereas Cyrillic is encoded as Unicode (2 bytes per letter). When the data was being entered into the original Word document, the student working on the material had to switch between English and Ukrainian keyboards on the computer. We wondered: could the student have forgotten to switch? We used several computer programs we wrote to check the English file names for Cyrillic and the Cyrillic categories for Latin letters. This was indeed where the problems were and, when we corrected the problem letters and the new XML file was uploaded the program worked.

The Ukrainian Folklore Sound Recordings website, which is available on the internet at: <http://www.artsrn.ualberta.ca/UkraineAudio/> has been on-line since 2006. In 2010 a major revision was made to the site: we added many sub-categories, placed the parts of rituals in chronological order, and added English translations of the categories. The addition of English translations is important because many North American members of the Ukrainian Diaspora can speak Ukrainian, having spoken it within the home, but cannot read or write the language, having had no formal training. My goal has always been to make my data accessible to as wide an audience as possible and adding English translations increased the number of potential users of my sound files many fold.

The current Ukrainian Folklore Sound Recordings website shows a first page which lists the major Topics, or metadata units, in both Ukrainian and English translation. If the user clicks on a topic such as: “весілля/The Wedding”, then the webpage with the first level of sub-categories under “весілля/The Wedding” appears. If one of sub-categories is picked, then the next levels appear. This is repeated until the required sub-sub----keyword is reached, for example:

весілля->обрядове вбрання->під час весілля->молодята

The Wedding->Ritual clothing->During the wedding->The bridal couple.

At this level the user will see the list of the relevant sound files themselves. In other words, those recordings which contain that final topic will be listed. The user then selects a sound file and a page displaying the interview information appears. This gives the information about the recording, such as the name of the respondent and the date and location of the recording. It also shows all of the indexed time points. At the bottom of the page a sound-player appears. On it, the time appropriate to the desired topic is highlighted. The user can then listen to the chosen section of the interview.

Being able to go directly to a sound file instead of relying on a transcription has many virtues in addition to saving time. When one can hear the actual recording, one

can get the nuances of intonation, inflexion, volume, and speed which are difficult to reproduce on paper. Also the context in which a particular topic is discussed can be retrieved and listened to by moving the cursor on the small sound-player to another part of the recording. Context, as many current studies have shown, can be crucial to understanding a text and its meaning [1]<sup>1</sup>. In addition, if the researcher looks through the Interview Index at the top of the page and notices an interesting topic that is different from the one originally searched for, then he or she can move the cursor of the sound-player on the screen until the appropriate time in the recording is shown – and listen to the new section.

Ukraine is not the only place where Ukrainians live. Canada has a sizeable Ukrainian Diaspora, one of the largest in the world. The first settlers, usually called the first wave of Ukrainian immigration, came 125 years ago and today there are approximately 1.2 million Canadians with some Ukrainian ancestry. Moving to Canada allowed me to add Canadian fieldwork to my work in Ukraine. Most of my work has been in the prairie provinces of Alberta and Saskatchewan, with some limited research in Manitoba and urban centres in the East. The prairie provinces are home to the highest concentration of Ukrainian Diaspora and have been the focus of my attention.

There have been 4 waves of Ukrainian migration. The first settlers came in the period from the late 1800s to the beginning of the First World War. They were primarily economic migrants, looking for a better life. They were working people, mostly farmers, who homesteaded and continued to farm in Canada. Some took jobs in construction: on the railroads, in mines, and clearing forests, hoping to earn enough money to buy their own land. Immigration was slowed by the war, but then resumed at war's end. The second, or interwar, wave of immigrants was similar to the first. The Second World War also affected immigration to Canada adversely. When a new wave of immigrants came after that war, they were different from their predecessors. Most were refugees who had been city dwellers and they came for political reasons. The fourth and most recent wave has been the one that started arriving after Ukraine declared independence.

The first two waves of immigrants are called the Pioneers. They settled all over Canada, but were particularly concentrated in the central and western provinces of Manitoba, Saskatchewan and Alberta. Ukrainian Pioneers were initially very poor. With hard work they prospered. One of the first things into which they invested their money was churches. They hired architects and woodworkers and donated countless hours of their own labor. They brought icons and crosses with them from Ukraine. They commissioned outstanding painters such as Peter Lipinsky and Theodore Baran. They decorated their churches with exquisite embroideries, fine woodwork, intricate stenciling on the walls, and carefully wrought chandeliers. The number of churches that they built is truly astounding.

The abundance of Ukrainian churches has, in a sense, been their downfall. When Ukrainians first arrived, they settled on quarter-sections, 160 acre plots of land. When farming was not mechanized, this worked well. However, as Ukrainian prosperity in-

---

<sup>1</sup> To write the article “How God Paired Men and Women: Stories and Religious Revival in Post-Soviet Ukraine” [1], I searched my database for religious stories about marital relations and then examined the context which prompted my respondents to volunteer these stories.

creased, as mechanization came to Canada, as roads improved, farmers bought up larger and larger tracks of land. Their sons and daughters moved to cities for education and for jobs. The rural population dwindled and is continuing to do so. There are simply not enough people to support all of the lovely little churches across the prairies. Churches that had one hundred and more members in the 1950s and 60s now have ten or less. The rural churches especially are no longer viable. In my fieldwork I have encountered many sad situations and one of the most telling was meeting a direct descendant of the Pylypiv family, one of the first families to come to Canada, outside of the decaying remnants of what had been his family church. The church was a ruin, but Pylypiv and his wife were still trying to save the cemetery, planting twelve pine trees for the twelve apostles and also to keep the neighboring marsh from encroaching on church burial grounds.

The Ukrainian churches of the Canadian prairies are a unique cultural phenomenon. The area stretching across Alberta, Saskatchewan and Manitoba is the largest colonization of post-Byzantine sacral culture in the New World. Beautiful little churches and some larger ones, topped with their emblematic onion domes, are a cherished part of the prairie landscape. The churches are beautiful inside as well as out. Many, the Bukovinian churches in particular, have stunning woodwork. Icons range from primitive ones painted on wood to sophisticated works executed by professional iconographers. As beautiful as these churches may be, as precious as they are to the people who attended and attend them, many, if not most, may soon be gone.

The importance of Ukrainian sacral culture and the threat of the imminent disappearance of Ukrainian churches on the Canadian prairies inspired the **Sanctuary Project**, an effort to document Ukrainian Canadian sacral heritage. We are a team of three faculty members from the University of Alberta. The team members: John-Paul Himka, Frances Swyripa and I, began this work 8 years ago. We dedicate a month every summer to traveling from church to church, documenting not only the exteriors, but also every item within each church and on its grounds. I conduct interviews with church members to capture a picture of church life. During the winter months, while we are teaching, we process our data and will soon open a photo database and an audio website. It is this audio database and its characteristics that have produced the next major advance in my documentation and presentation of Ukrainian culture and folklore.

When we first started processing the sound files that I had recorded as part of the Sanctuary Project, we used the same system as the one used for the sound files recorded in Ukraine and described above. The materials processed so far are available at: <http://www.artsrn.ualberta.ca/EnglishAudio/>. I was also able to do some work among the Ukrainians of Kazakhstan. With support from the Kule Chair Endowment and with the guidance of Alevtina Cvetkova of the Toraigyrov State University in Pavlodar, I was able to collect data about ritual practices in the villages of northeastern Kazakhstan. This area was interesting to me because Ukrainians settled there at almost the same time as when they first started coming to Canada. The Stolypin reforms brought a sizeable migration from Ukraine to the Pavlodar area, as villages with names such as Krasilovka, Sosnovka, and Khmelnytskoe attest. The soundfiles from this work were also indexed according to the system already in place and can be listened to at <http://www.artsrn.ualberta.ca/KazakhstanAudio/>.

The success and the importance of the Sanctuary Project brought it increasing attention and we were able to generate grant support from the Killam Foundation, Alberta Heritage Foundation, and Saskatchewan Heritage, among others. The greatest impact on our work came with a KIAS (Kule Institute for Advance Study) Cluster Grant. This grant supported fieldwork and paid for graduate student assistants. It also provided for support from Alberta Libraries. In agreement with the library system, our work will become part of the Peel's Prairie Provinces collection and be maintained in perpetuity by the library. Working with library staff, we decided to use a different coding system for our data. The images generated by my colleagues Himka and Swyripa will become part of one database, a preliminary version of which has already been built. My sound files will become part of another database and that work will be discussed here.

The library staff shared my goal of making my data widely available. With that goal in mind, Sharon Farnel and other library workers suggested converting our in-house indexing system to annotation using the metadata in the American Folklore Society Ethnographic Thesaurus: <http://id.loc.gov/vocabulary/ethnographicTerms.html>. This thesaurus is the international standard and using it would make my recordings easily accessible around the world; it would allow my work to reach both scholars of Ukrainian folklore and researchers interested in other folklores and other world areas.

Re-coding my sound files, even the ones already indexed, proved no easy task. The categories in the Thesaurus were so different from the metadata that we had generated in-house that we decided to start by coding those sound files that had not been processed, planning to return to the earlier files and to re-code them at a later date. So far, using the new, Ethnographic Thesaurus system, we have processed the recordings made in the years 2013, 2014, and 2015 or 79 hours of audio.

The Ethnographic Thesaurus was produced by the American Folklore Society in the United States. As a result, it is United States-centric not only in terms of the metadata units selected, but also in terms of the arrangement of categories. For our work, the Thesaurus presents two problems. One is that the categories of interest to Ukrainians are often embedded many layers down into the coding system. The other is that terms typically used by Ukrainian scholars and Ukrainians themselves either do not appear in the Thesaurus at all or are coded using terminology that is oriented toward Western European and American usage: it uses terms for holidays and other ritual categories that Ukrainians do not use and might not think of. There are also some structural problems with the Thesaurus which will be mentioned briefly after the Ukrainian-specific discussion.

The problem of deep embedding can be seen in the following two examples. Churches are extremely important to Canadian Ukrainians and were a central focus of our work. Yet, in the Thesaurus, the term "*church*" appears as the last in a long line of categories, namely: material culture, objects, architecture, buildings, religious buildings, churches. It is the last in a string of six terms. Similarly, Easter is a central holiday for Ukrainian Canadians and Ukrainians in general, yet it is eighth in a string of terms which are: ritual, calendar rites, holidays, religious holidays, Lent, Holy Week, Easter. While the search engine will take the user directly to the category desired, the category structure does convey relational levels and may make Ukrainian material seem less important.



An inverse problem is the lack of sub-categories that are important for Ukrainian material. Food is a central part of all rituals and some foods, like the *korovai* that is part of the wedding, the *paska* and *babka* that are part of Easter, and the *kolach* which appears both at Christmas and as part of funeral rites and commemoration of the deceased are specifically Ukrainian and central to Ukrainian tradition and lore. None of these terms can be found in the Thesaurus as it is currently structured. More importantly, the lowest category in the food metadata set is ritual meals. As already noted, ritual meals are part of every Ukrainian ritual and a user searching under that category would get 193 hits in just the sound file data that we have processed so far. Not only would that person not be able to find out whether the *korovai* is used as part of weddings in a certain area of Canada or in a certain time period, he or she would not be able to distinguish between the banquet that follows a wedding and the fundraising meal that comes as part of Malanka. The user would have two options. One is to click on each of the 193 options that the search for ritual meal yields and check each to see what kind of meal is being discussed and to which ritual the meal belongs. The other option is to search by ritual rather than by meal. Because the Ethnographic Thesaurus is United States-centric, the arrangement of categories is influenced by current trends in American scholarship. Foodways is a popular subdivision of folklore and, for a foodways specialist, whether a meal is a ritual meal or belongs to another food category would be more important than which ritual a meal accompanies. For Ukrainianists, the opposite might be true. Western scholarly interests are also evident in the Thesaurus distinction between icons and icon painting. Icons belong under ritual and ceremonial objects and icon painting belongs under art. Such a distinction is not useful for our material and is governed by the fact that American folklorists would likely be interested in either art or ritual, but not the combination of the two.

The problem of absence of Ukrainian terminology is a serious one. Saints' days is a category in the Ethnographic Thesaurus, but many saints important to Ukrainians are not listed. These include saints after whom prairie churches are named such as Saints Peter and Paul and Saint Nicholas. In the prairies, one of the most important events is precisely the service held on the day of the saint after whom a church is named. It is called *Khram* among the Orthodox and *Praznyk* among Eastern rite Catholics. The terms *Khram* and *Praznyk* are not part of the Thesaurus either and cannot be used as an alternative for Saints' Days. *Churching of the mother*, a ritual that is sometimes combined with baptism, but is sometimes performed on a separate occasion, does not appear in the Thesaurus. The Thesaurus does list the category "baptism" but no sub-categories follow. There are numerous other examples. Canadian Ukrainians almost all remember going as children to neighbors' and relatives' homes to perform the practice of *Siiaty*, a ritual which involves throwing grain and reciting a little verse that wishes the home being visited prosperity and good fortune. This event does not exist as a Thesaurus category and the closest category that we can find is New Year.

Some important events are coded under terms which Ukrainians do not use and are not likely to search for when using our database. In other words, there are certain calendar holidays that are known by one name among Westerners and by another name among Canadian Ukrainians. For example, Ukrainians in the prairies use the term *Pussy Willow Sunday* for the Sunday before Easter and have little rhymes and sayings

connected to the pussy willows that characterize that day. The term “Pussy Willow Sunday” does not exist in the Ethnographic Thesaurus and the Sunday before Easter is designated as Palm Sunday and that alone. Some Eastern rite Catholic churches use the term Palm Sunday and which term a respondent chooses to use may well be indicative of religious affiliation. Thus, for our research, the distinction between Palm Sunday and Pussy Willow Sunday is important. *Trinity Sunday*, a summer holiday important to Ukrainians and one after which a number of prairie churches are named, is not a term listed in the Ethnographic Thesaurus. The Thesaurus term is Pentecost and no synonyms for this holiday name are given. Ukrainians will often refer to this day as Green Holidays, but Pentecost is not a term that they will readily recognize. The water-blessing event that comes twelve days after Christmas and which Ukrainians call *Yordan* is coded as Epiphany in the Thesaurus. If Ukrainians do use a term other than *Yordan* for this holiday, they will call it *Theophany*. The Thesaurus term “Epiphany” is rarely used by Ukrainians and, while searching under Epiphany will guide non-Ukrainians to Ukrainian material, Ukrainians who search for this holiday under the designation “Yordan,” the term with which they are familiar may get the impression that information about this event is not available in our database.

Because we are committed to using the Ethnographic Thesaurus and thus bringing worldwide attention to Ukrainian material, we have adopted a two-pronged solution to the problems it presents. One has been to add Ukrainian-specific categories: to date we have identified and added 24 such categories. By this we mean that we have added all of the Ukrainians-specific terms listed above and more to our set of metadata. We do have the metadata terms *Yordan* and *Pussy Willow Sunday*, and *Trinity Sunday*, and *Siiaty*, and all of the saints after whom churches are named and who are honoured on their respective *Khram* or *Praznyk*. For terms like *Pussy Willow Sunday*, we have used the more Western designation of *Palm Sunday* if that is the terms used by the speaker on the recording. If the speaker chose to refer to the holiday in question as *Pussy Willow Sunday*, then that is the designation used.

In the material processed so far we have 2223 indexed sections and 193 categories. Of these, the 24 categories that are uniquely Ukrainian account for 349 indexed sections. The remaining materials (those which do correspond to the American Folklore Society Ethnographic Thesaurus) comprise 1874 indexed sections in 169 categories. We are confident that the 24 new categories we have identified are valid and do indeed reflect Ukrainian Canadian ritual practice. Our confidence is based on the fact that our work shows a high degree of recurrence. Of these new categories: 13 are found in 10 or more sections and 9 are found in 20 or more sections. *Yordan*, for example, is talked about in interviews 46 times.

Identifying these 24 Ukrainian Canadian terms means that we have the potential of expanding the Ethnographic Thesaurus by 24 metadata units. We will be holding a roundtable at the meeting of ASEES, the Association for Slavic, East European and Eurasian Studies scheduled for Washington DC in November, 2016. At the roundtable we will discuss Slavic-specific ethnographic categories and draft a statement to the American Folklore Society, proposing expansion of the Ethnographic Thesaurus to include the Ukrainian and other Slavic terminology that we have identified. The Ethnographic Thesaurus has added categories from other cultures that practice their traditions

in North America. One example is the Hindu calendar ritual of Diwali. We will work for the inclusion of Ukrainian-specific and more generally Slavic categories in this international metadata system. We have been working with the Ethnographic Thesaurus for slightly over one year and we have observed that it is in flux. Categories do change, as do the relationships between metadata units. This appears to be a particularly opportune time to suggest new Ukrainian-specific categories for inclusion.

Our other tactic will be to include two search pathways on the entry page of our database. One will be a pull-down menu which will go by the Ethnographic Thesaurus and should attract users not familiar with Ukrainian culture who wish to learn more about Ukrainian folklore and ritual or who wish to look for Ukrainian material to compare to the culture which they are researching. It will serve those who are not Ukrainian specialists and want to parallel Ukrainian phenomena to the geographical and cultural area of their research. The tree on which this search portal will be based can be seen at our preliminary website <http://www.artsmn.ualberta.ca/orodrigu/tree/>. The other pull-down menu on the opening page will have far fewer metadata terms and will be geared to Ukrainian users. Terms like Yordan and Trinity Sunday and churching of the mother will come up immediately and not be imbedded in the great number of categories that characterize the Ethnographic Thesaurus. I am also considering writing an introductory paragraph about Ukrainian Canadian culture. We will need to create the two pull-down menus and to test them to see what additional guides, such as an introductory paragraph, will be needed.

Coding Ukrainian Canadian data according to a metadata system that is not geared to this type of material is a challenge. It is a challenge worth accepting. From the beginning of my ethnographic work I have sought to make Ukrainian culture as widely known as possible. My students at the University of Virginia had no Ukrainian background. Having them work with my Ukrainian website in order to learn the principles of folklore scholarship helped promote awareness of Ukraine and Ukrainian culture. The database with the audio recordings made in Ukraine gave the public access to Ukrainian language material. It has been used primarily by researchers, but North Americans with a knowledge of spoken Ukrainian, including those with no knowledge of the written language, have also benefitted from the site. Some of my Ukrainian Canadian English language sound files are already available using our old, in-house set of search terms. The new database, built according to the Ethnographic Thesaurus, should be available shortly and it should reach the broadest possible audience. We are building this site in such a manner that it will also provide easy access to users from the Ukrainian community. They will be served by a search bar geared specifically to Ukrainian interests. Looking beyond our own database and working to have an impact on the international scale, we will seek to expand the Ethnographic Thesaurus to include Ukrainian categories, thus bringing awareness of Ukrainian culture to all people who use the Thesaurus for whatever purpose, including coding their own material. Even those people who never use our database but who are folklorists turning to the Thesaurus to categorize their own material should become aware of Ukrainian folklore through its inclusion in this internationally recognized system.

## References

1. *Kononenko Natalie*. How God Paired Men and Women: Stories and Religious Revival in Post-Soviet Ukraine // *Canadian-American Slavic Studies*. – 2010. – Vol. 44. – № 1–2. – P. 105–133.

2. *Kononenko Natalie*. *Ukrainian Minstrels: And the Blind Shall Sing*. – Armonk, New York, London: M.E. Sharpe, 1998. – XVI, 360 p., ill.

3. *Брицина Олександра, Головаха Інна*. Прозовий фольклор села Плоске на Чернігівщині: Тексти та розвідки. – Київ: Інститут мистецтвознавства, фольклористики та етнології імені М. Т. Рильського НАН України, 2004. – 584 с.

Стаття надійшла до редколегії 09.10.2016

Прийнята до друку 25.10.2016

## СТВОРЕННЯ БАЗИ ДАНИХ УКРАЇНСЬКОГО ФОЛЬКЛОРУ: ПРЕДСТАВЛЕННЯ ІНФОРМАЦІЇ МІЖНАРОДНІЙ АУДИТОРІЇ<sup>1</sup>

Наталія КОНОНЕНКО

*Peter & Doris Kule Centre for Ukrainian & Canadian Folklore,  
Modern Languages and Cultural Studies, Faculty of Arts, University of Alberta  
441-C Old Arts & Convocation Hall,  
tel.: 780-492-6810, e-mail: natalie.kononenko@ualberta.ca*

Польове дослідження фольклору нагромаджує великі масиви даних. Завдання вченого полягає в опрацюванні цих даних і наданні доступу до них максимальній кількості користувачів. Робота над вебсайтами з українським фольклором розпочалася в 2000 р. Вона допомогла краще зрозуміти відмінності між потребами дослідника, студентів і загальних користувачів задля ефективного користування сайтом. Презентації аудіоматеріалів в інтернеті особливо доцільні, оскільки добре передають особливості мовлення. Проблематичним видається створення баз даних із фольклорними аудіофайлами, які б відповідали потребам як дослідників, так і широкого загалу. Вебсайт із записами українського фольклору (*Ukrainian Folklore Sound Recordings*) вперше скеровував користувача безпосередньо до запису на вибрану тему. Сучасна практика кодування використовує метадані Міжнародного етнографічного тезауруса (*The International Ethnographic Thesaurus*), а база даних проекту “Святині” (“*Sanctuary*”) зі звуковими файлами українсько-канадських записів працює за цією системою. Робота з кодування, яка проводиться в рамках проекту “Святині”, значно збагатила Етнографічний тезаурус українським матеріалом.

*Ключові слова*: польове дослідження фольклору, цифрові технології, український фольклор у Канаді, вебсайт *Ukrainian Folklore Sound Recordings*, *International Ethnographic Thesaurus*.

<sup>1</sup> Переклад українською мовою Софії Кірей. Переклад авторизований. – *Примітка упорядника*.

Я розпочала польову роботу зі збирання етнографічного матеріалу в Україні в 1998 році. За професією я фольклорист-етнограф, і польова робота є важливою частиною моєї діяльності. Але займатися записами польового матеріалу в Україні не було можливо: статус України у складі Радянського Союзу обмежував можливість подорожувати країною і супроводжувався контролем за доступом до інформації взагалі й особливо до українського села. Лише після розпаду Радянського Союзу, коли Україна стала незалежною державою, польова робота стала можливою, і особливо для таких іноземців, як я. Я працювала в Україні і раніше. У радянський період проводила дослідження в архівах, де готувала до публікації книгу про українське кобзарство – мистецьку традицію, яка майже зникла до того моменту, коли я збирала ці матеріали [2]. У радянські часи можна було досліджувати лише народні традиції минулого, оскільки вивчати сучасні явища під радянським контролем було майже неможливо. З проголошенням незалежності України я отримала змогу їздити у сільські місцевості й збирати інформацію про тамтешнє життя. Я мала можливість брати інтерв'ю у людей про такі обряди, як весілля та хрестини, і спостерігати за цими діями. Захоплення документуванням раніше недоступних матеріалів заохотило мене поділитися ними зі своїми студентами. Це, а також моє зацікавлення цифровими технологіями, викликало в мене бажання створити веб-сайт українського фольклору.

У часі цих польових студій я працювала у Вірджинському університеті (США), що був лідером у запровадженні комп'ютерних технологій у гуманітарній сфері. У 2001 році завдяки фінансуванню та технічній підтримці Університету я створила проект **“Українське село”** (“The Ukrainian Village Project”). Він складався з бази фотографій, які можна було використовувати з дослідницькою метою, і веб-сайту – для використання під час моїх занять і для надання доступу до цієї інформації широкому загалу. Спочатку у мене була мета створити повну базу даних. Я була захоплена можливостями комп'ютерних технологій і гадала, що за допомогою своєї бази даних зможу передати широкій аудиторії (і своїм студентам, і неакадемічним користувачам мого сайту) всі нюанси фольклорних явищ. Я сподівалася, що мого веб-сайту буде достатньо для того, щоб зробити мою роботу широкодоступною. Створення бази даних багато чому мене навчило. Моїм початковим припущенням було те, що надання інформації онлайн дасть мені змогу відобразити складність фольклору краще, ніж я могла би це зробити шляхом написання статті чи монографії. Виявилось – зовсім навпаки. Взаємозв'язок між сферами фольклору є багатоаспектним. Матеріальні об'єкти, такі як *рушники*, використовують для прикрашання і захисту домівки, але вони також є обрядовими предметами, що використовуються під час весілля і похорону. Ліричні пісні виражають емоції. Історичні пісні теж виражають емоції, але вони також фіксують і описують важливі історичні події. Коли я створила свою першу базу даних, ту, що складалася з фотографій і описів, я намагалася охопити комплексність фольклору, але мені не вдалося передати сутність української культури і традицій моїм студентам. Ця база даних була корисною для мене, тому що я вже була ознайомлена з українським фольклором і його взаємозв'язками. Але використавши свої бази даних в аудиторії, я зрозуміла, що для людей, яким ця тема незнайома, інформація на моєму сайті була занадто складною, щоб її можна було

застосовувати ефективно. Я усвідомила, що студентів, які не мають попередньої підготовки і не знають української культури, треба ознайомлювати з матеріалами в простіший спосіб. Складні взаємозв'язки, які я могла передати в цифровій формі, були за межами сприйняття студентів, які робили свої перші кроки в цій галузі. Щоб ефективно навчати своїх студентів, мені варто було розпочати з чогось простого і потім перейти до складніших явищ. У підсумку я залишила свою базу даних для власних досліджень, а для своїх студентів розробила веб-сайт з вибраними зображеннями і простим текстом, який представляв вибірку інформацію про український фольклор. Цей сайт став відомим під назвою **“The Ukrainian Village Project”** (“Українське село”) і був дійсно корисним інструментом для викладання. Він також отримав добрі відгуки від загальної публіки. Сайт постійно оновлюється. Тепер він називається **“Folklore Ukraine”** і доступний за посиланням: <http://www.artsrn.ualberta.ca/folkloreukraine/>.

Основне оновлення сайту “Folklore Ukraine” відбулося в Університеті Альберти (Канада). Я долучилася до факультету цієї установи в 2004 році і за підтримки технічного персоналу Інформаційного центру при Кафедрі гуманітарних наук (The Arts Resource Centre) та за фінансової підтримки Центру викладання і навчання (Centre for Teaching and Learning) розпочала роботу над веб-сайтом “Folklore Ukraine” та створила інші веб-сайти. Створення загальнодоступних веб-сайтів корисне для спільноти і є чудовим способом презентувати свою роботу. Але я прагнула більшого. Я постійно думала над комп'ютеризованим пошуком даних. Мені пощастило, що моєю аспіранткою у той час була Світлана Кухаренко. Ми зі Світланою почали індексувати численні звукові файли, які я записала в Україні. Моєю метою було зробити всі мої звукові записи доступними для пошуку. Процес був складним. Більшість записів було зроблено на касетах і їх необхідно було оцифрувати. Значення процесу оцифрування – очевидне. Оскільки касети є предметами матеріальними, а не віртуальними, то щоб знайти певний уривок запису, потрібно перемотувати плівки вперед і назад, аж поки потрібний фрагмент не знайдено. Переміщення курсора в межах цифрового запису не потребує жодних маніпуляцій з плівкою, а отже, в разі збільшує швидкість пошуку. Але все ж процес пошуку матеріалу, навіть у цифровому записі, може бути клопітким. До того ж, коли я розпочала спроби індексувати свої звукові файли і робити їх доступними для пошуку, я збрала більше ста годин звукозаписів. Оскільки мої записи були спочатку на касетах, то їх було розділено на блоки тривалістю у сорок п'ять хвилин, і таких блоків виявилось багато. Пошук серед (мається на увазі всередині кожного з цих файлів. – *Примітка перекладача*) близько двохста файлів забрав чимало часу. Мені допомагали мої детальні анотації до даних, але, беручи до уваги обсяг матеріалу в моєму розпорядженні, для швидкого пошуку їх виявилось недостатньо. В анотаціях я зазначала дату, місце і час запису, а також людей, у яких брала інтерв'ю, та вчених, присутніх під час нього. Я занотувувала зміст записів, а саме: вказувала ті плівки, де респондент говорив чи про шлюбні обряди, чи про похоронні, чи розповідав релігійні легенди або казки. Всі мої нотатки були оцифровані. Однак вони могли лише вказати мені відповідний сорокап'ятихвилинний аудіофрагмент. Вони не показували, де конкретно у цих сорокап'яти хвилинах був матеріал, який я хотіла знайти.

Один зі способів зробити пошук польових матеріалів простішим і швидшим – це виконати транскрипцію всіх звукових файлів. В Україні обробку звукових записів було зроблено саме так. Транскрипція означає, що людина слухає запис і нотує все, що сказано. Транскрипція аудіозапису є дуже корисною: створюється текстовий файл, у якому, особливо в цифровій формі, можна легко знайти необхідний фрагмент за допомогою загальнодоступної пошукової системи, яка подається в комплекті з MSWord чи іншим програмним забезпеченням для обробки текстів. Можливість внутрішнього пошуку тексту значно скорочує час, необхідний для пошуку певного фрагмента. Однак транскрипція має два недоліки. Перший – фінансовий. А саме, на Заході можливість знайти й оплатити роботу людей, які б виконали транскрипцію ста годин записів, була поза обговоренням. Іншим недоліком є втрата автентичного звучання і всіх акустичних вказівок до значень, які воно із собою несе. Фольклор є усним. Фольклор – це мистецтво. Значна частина з того, що фольклор передає, виражається мистецькими засобами. Наративи, наприклад, характеризуються голосовими модуляціями: змінами в тоні, гучності, швидкості і так далі. Відтворити їх на папері складно. Деякі спроби в цьому напрямі все ж були зроблені, що відображено в книзі “Прозовий фольклор села Плоске”, яку опублікувала моя колега Олександра Бріцина [3]. Тут підвищений темп показаний жирним шрифтом; акцент зазначений курсивом і так далі. Такі конкретизовані транскрипції, хоча корисні, потребують ще більше часу, ніж звичайні. Це також збільшує витрати на публікацію – проблема, яка стає все актуальнішою на Заході. Крім того, вони не повністю відображають експресивний потенціал оригіналу, який можна було б почути. На Заході, поза межами фольклористики, було зроблено декілька повних транскрипцій масштабних групових проєктів. Одним із прикладів є Британський національний корпус (British National Corpus, BNC). Він налічує сто мільйонів слів з можливістю внутрішнього пошуку в колекції усних та письмових зразків сучасної британської англійської мови. Проте навіть ця широкомасштабна робота має свої проблеми, і багато розділів не було затранскрибовано, а лише позначено як “незрозумілі” (“unclear”).

Бажаною формою підготовки польових записів для дослідження є безпосередня робота зі звуком і анування звукових файлів. Я і мої колеги з технічної підтримки та моя аспірантка Світлана Кухаренко вирішили індексувати мої звукові файли і використати цей індекс як метадані; на нашим задумом, це мало би дозволити вченим перейти безпосередньо до місця в оригінальному і автентичному звуковому записі, де обговорюється потрібний зразок інформації.

Індексування полягає у зазначенні всіх тем, які обговорено в записі, і точного часового моменту, коли відбувалося обговорення. Індексування потребує шаблону – “основного шаблону індексування інтерв’ю” (“Master Interview Index Form”). Цей шаблон повинен бути адаптований до видів інформації, наявної в інтерв’ю. Мене цікавлять такі обряди, як весілля, похорони і народження; відповідно, вони були позначені як основні категорії. У межах основних категорій було визначено підкатегорії – етапи тих обрядів, які були позначені як основні категорії. Процес індексування потребує прослуховування звукового запису і зазначення часу, коли згадується одна з цих категорій або підкатегорій (за допомогою ключових слів).

Ключові слова і їхні часові позначки з інтерв'ю вносять у “форму індексування інтерв'ю” (“Interview Index Form”). Верхня частина такої форми містить назву запису і його дату, місце і осіб, у яких брали інтерв'ю. Після створення шаблон можна змінювати в процесі роботи, і індексування відбувається відносно швидко. Вже немає потреби повертатися уперед і назад та прослуховувати запис кілька разів, як під час виконання транскрипції. Як правило, індексувати матеріали можна з першого разу, в той час, коли матеріал опрацьовується, особливо якщо той, хто індексує, знайомий з шаблоном. Для кожного звукового файлу у процесі індексування надається список шаблонних категорій, які присутні в цьому записі, та часових позначок, коли ці теми звучать у запису. У такий спосіб було опрацьовано оригінальну добірку зі 150 годин mp3 записів, а загальний індекс зайняв 152 сторінки текстового документу Word. Ось приклад трьох послідовних позицій, індексованих у записі ploske2000s.mp3:

04.04–06.25	весілля – чесна/нечесна молода;
06.26–06.48; 12.13–12.58	весілля – сніданок для молодих;
14.56–16.26	весілля – ділять коровай/дарування.

Цей документ виявився дуже корисним для мене, адже за його допомогою можна було знайти потрібну тему. Пошук вказував, у яких mp3-файлах і в які часові проміжки ця тема звучала. Mp3-файли можна було прослуховувати, починаючи від зазначеної часової позначки. З просуванням моєї роботи мені хотілося автоматизувати процес пошуку: я бажала мати “зв'язок” (“link”) між різними темами та конкретними фрагментами відповідного звукового файлу. З наведеного вище прикладу потрібно поєднати три речі, щоб такий автоматичний пошук став можливим. Ці три речі є: назва запису (1), часова позначка (2) і тема (3).

- 1) ploske2000s.mp3;
- 2) 06.26–06.48; 12.13–12.58;
- 3) весілля – сніданок для молодих.

Спочатку ми планували розділити оригінальний mp3-файл на невеликі сегменти: по одному для кожної представленої теми (наприклад: “ploske2000s 06.26–06.48”), і поєднати кожен сегмент з його темою. У такий спосіб, коли потрібну тему буде знайдено, програма відтворюватиме відповідний невеликий сегмент. Такий підхід має три проблеми: по-перше, довелося б вирізати близько 5 000 сегментів; по-друге, якби трапилася невелика помилка в індексуванні часових позначок, сегмент був би змарнований; і по-третє, контекст сегмента в цілому записі був би втрачений. На щастя, нам не довелося нічого вирізати, тому що програміст Ерік Джанг (Eric Yue Zhang), який працював у Дослідницькому відділі текстового аналізу (Text Analysis Portal for Research) Університету Альберти, написав пошукову програму і показав нам, як “пов'язувати” теми зі звуковими сегментами без вирізання. Дані, за допомогою яких програма виконує пошук, наявні в XML-файлі, який ми склали на основі інформації з вихідного текстового документу (Word) “головної форми індексування інтерв'ю”.

На жаль, коли створений XML-файл було завантажено на сервер, де знаходилася пошукова програма, ця програма не запрацювала. Причиною була помилка в коді XML. Незважаючи на те, що ми описали його кількома словами, процес конвертування даних зі 152-сторінкового документу Word у файл XML, що містив



23 129 рядків, не був простим. Щоб перетворити його у формат XML, ми використовували кілька власноруч написаних комп'ютерних програм, а також неодноразово застосовували функцію документа Word "знайти і замінити". Браузер Mozilla Firefox, який показує помилки форматування в файлах XML при спробі їх відкрити, на жаль, не бачив жодних очевидних помилок форматування. Причиною ж виявилася візуальна схожість між деякими літерами латинського алфавіту і кирилиці та відмінності в їх комп'ютерному кодуванні. Хоча літери **a, c, o** та ін. виглядають ідентично англійською мовою і кирилицею, англійські кодуються як ASCII (один байт на одну букву), натомість кирилиця кодується як Unicode (2 байти на букву). Коли дані вводили в початковий документ Word, студентів, який працював з матеріалами, доводилося перемикаати англійську та українську клавіатури на комп'ютері. Ми подумали: чи міг студент забути перемкнути клавіатуру? Ми використали кілька власноруч написаних комп'ютерних програм, щоб перевірити англійські назви файлів на наявність кирилиці, а кириличні назви категорій – на наявність латинських букв. Змішування кирилических та латинських літер і було джерелом проблеми. Тож коли ми виправили проблемні букви і завантажили новий XML-файл, програма запрацювала!

Веб-сайт звукозаписів українського фольклору, який доступний в Інтернеті за адресою: <http://www.artsm.ualberta.ca/UkraineAudio/>, функціонує в режимі онлайн з 2006 року. У 2010 році у сайт було внесено низку змін: ми додали багато підкатегорій, розмістили підкатегорії обрядів у хронологічному порядку і додали англійські переклади. Додавання англійського перекладу є важливим, адже багато хто з північноамериканської української діаспори може говорити українською мовою або говорить нею вдома, але не може читати або писати українською, оскільки вони не мали формального навчання. Моєю метою завжди було зробити мої дані максимально доступними, тому додавання англійських перекладів значно збільшило кількість потенційних користувачів моїх звукових файлів.

Теперішній веб-сайт звукозаписів українського фольклору показує головну сторінку, на якій містяться основні теми (або одиниці метаданих) як українською, так і англійською мовами. Коли користувач натискає на тему, наприклад, "весілля/The Wedding," то під нею з'являється веб-сторінка з першим рівнем підкатегорій. Якщо обрати одну з підкатегорій, з'являються наступні рівні підкатегорій. Це повторюється доти, поки користувач не дійде до необхідного під-під-ключового слова, наприклад:

весілля->обрядове вбрання->під час весілля->молодята

The Wedding->Ritual clothing->During the wedding->The bridal couple.

На цьому рівні користувач сам побачить список відповідних звукових файлів. Іншими словами, ті записи, які містять потрібну тему, будуть перераховані. Потім користувач вибере звуковий файл і з'явиться сторінка, що висвітлює інформацію з інтерв'ю. Вона відображає дані про запис, наприклад: ім'я респондента, дату і місце запису, а також всі індексовані позначки часу. Унизу сторінки з'явиться звуковий програвач, на якому позначено часові проміжки, що відповідають темі, яку шукали. Так користувач може прослухати вибраний фрагмент інтерв'ю.

Можливість перейти безпосередньо до звукового файлу, замість того, щоб покладатися на транскрипцію, окрім економії часу, має багато інших переваг. Коли є

можливість прослухати справжній звукозапис, можна вловити нюанси інтонації, словозміни, гучності і швидкості, які важко відтворити на папері. Також можна виокремити і прослухати контекст, у якому обговорюється конкретна тема, шляхом переміщення курсора до іншої частини запису на звуковому програвачі. Як показало багато сучасних досліджень, контекст може мати вирішальне значення для розуміння тексту і його значення<sup>1</sup>. Крім того, якщо дослідник переглядає індекс інтерв'ю зверху сторінки і зауважує цікаву тему, яка відрізняється від тої, яку він спочатку шукав, він може переміщати курсор звукового плеєра на екрані доти, поки не з'явиться відповідний часовий фрагмент у записі, і потім слухати новий розділ.

Україна – це не єдине місце, де живуть українці. У Канаді є велика українська діаспора, одна з найбільших у світі. Перші поселенці, яких прийнято називати першою хвилею української імміграції, прибули сюди 125 років тому, а сьогодні в Канаді проживає близько 1,2 мільйона канадців українського походження. Переїзд до Канади дав мені змогу долучити до моїх українських польових досліджень канадські матеріали. Більша частина моєї польової роботи відбувається в преріях Альберти та Саскачевану, з деякими дослідженнями в Манітобі та містах на Сході. Я зосереджую свою увагу на цих прерійних провінціях, де сконцетрована найбільша українська діаспора.

Усього було чотири хвилі переселення українців до Канади. Перші поселенці прибули в період з кінця XIX століття до початку Першої світової війни. Це переважно були економічні мігранти, які прибували в пошуках кращого життя: люди з робочого класу, здебільшого селяни/землероби, які оселилися і продовжували займатися сільським господарством у Канаді. Деякі з них почали працювати на будівництві залізниць, шахт, а також займатися вирубкою лісів з надією заробити достатньо коштів, щоб придбати власну ділянку землі. Імміграція була сповільнена війною, але потім відновилася після її завершення. Друга, міжвоєнна, хвиля імміграції була схожа на першу. Друга світова війна також негативно вплинула на еміграцію до Канади. Коли після цієї війни прибула нова хвиля іммігрантів, вони відрізнялися від своїх попередників. Більшість із них були біженцями – жителями міста, які емігрували з політичних причин. Четверта, нещодавня хвиля, почала прибувати після проголошення незалежності України.

Перші дві хвилі іммігрантів називають Піонерами. Вони розселилися по всій Канаді, але найбільша кількість переселенців прибула до центральних і західних провінцій Манітоби, Саскачевану та Альберти. Українські Піонери спочатку були дуже бідними, але завдяки своїй тяжкій праці досягли успіху. Першим, у що українці почали вкладати свої кошти, були церкви. Вони наймали архітекторів і теслярів і самі віддавали цьому чимало годин праці. Українські іммігранти привезли з України ікони і хрести. Вони замовляли ікони у видатних художників, таких, як Петро Липінський (Peter Lipinsky) і Теодор Баран (Theodore Baran). Українці при-

<sup>1</sup> Для написання статті “How God Paired Men and Women: Stories and Religious Revival in Post-Soviet Ukraine” (“Як Бог людей парував: легенди та релігійне відродження у пост-радянській Україні”) [1], я шукала у своїй базі даних релігійні історії про сімейні стосунки, а потім вивчала контекст, який спонукав моїх респондентів розповідати ці історії.

крашали свої храми вишуканими вишивками, витонченими виробами з дерева, вигадливими розписами на стінах, які виконували за допомогою трафаретів, і філігранно кованими люстрами. Кількість церков, які вони побудували, справді вражає.

Велика кількість українських церков стала в певному сенсі причиною їхнього занепаду. Коли українці прибули вперше, вони селилися на ділянках площею 160 акрів землі [приблизно 64,7 гектара. – *Примітка перекладача*]. І поки сільське господарство не було механізованим, цієї землі вистачало. Однак зі зростанням добробуту українців, з приходом механізації до Канади, з покращенням доріг, ці перші іммігранти почали скуповувати все більші і більші ділянки землі. Їхні діти переїжджали до міст, щоб здобути освіту і знайти роботу. Сільське населення зменшилося і продовжує зменшуватися. Тож сьогодні недостатньо людей, які б підтримували всі ці чарівні маленькі церкви у преріях. Церкви, що в 1950–1960-х роках мали сто або й більше членів, тепер налічують десять, а то й менше осіб. Особливо занепадають сільські церкви. Під час польових досліджень я мала справу з багатьма прикрими ситуаціями, і однією з найбільш показових була моя зустріч з прямим нащадком сім'ї Пилипових (що однією з перших приїхала до Канади) біля залишків того, що колись було його родинною церквою. Церква була в руїнах, але Пилипів з дружиною все ще намагалися зберегти цвинтар, висаджуючи дванадцять сосен для дванадцяти апостолів; а також вберегти церковні поховання від наступу прилеглого болота.

Українські церкви канадських прерій є унікальним культурним явищем. Цей ареал, що простягається через Альберту, Саскачеван і Манітобу, є місцем найбільшої колонізації пост-візантійської сакральної культури в Новому Світі. Красиві маленькі церкви і кілька великих, увінчаних символічними куполами, є невід'ємною частиною степових краєвидів. Церкви прекрасні як усередині, так і ззовні. Багато з них, а особливо буковинські церкви, мають прекрасні вироби з дерева. Ікони варіюються від примітивних, намальованих на дереві, до вишуканих, виконаних професійними іконописцями. Однак якими б ці церкви не були красивими чи цінними для людей, що їх відвідували і продовжують відвідувати, багато храмів, якщо не більшість, незабаром можуть зникнути.

Важливість української сакральної культури і загроза неминучого зникнення українських церков у канадських преріях стали стимулом для створення проекту **“Sanctuary”** (“Святині”) – спроби задокументувати українську сакральну спадщину Канади. Ми – команда з трьох співробітників факультету Університету Альберти: я, Джон-Пол Хімка (John-Paul Himka) та Френсіс Свиріпа (Frances Swygi-ra) – почали цю роботу ще 8 років тому. Щоліта ми проводимо місяць у преріях, подорожуючи від церкви до церкви, і задокументуємо не тільки екстер'єр, але й кожну деталь у церквах і на територіях довкола них. Я проводжу інтерв'ю з членами церкви, щоб охопити повну картину церковного життя. Протягом зимових місяців, читаючи лекції в університеті, ми опрацьовуємо зібрану інформацію і незабаром відкриємо базу даних фотографій та аудіо-веб-сайт. Саме ця звукова база даних та її характеристики привели мене до чергового поступу в моєму документуванні та представленні української культури і фольклору. Коли ми почали опрацьовувати звукові файли, які я записала в межах проекту “Sanctuary”, ми ви-

користували ту ж систему, яку застосовували для звукових файлів, записаних в Україні (див. вище). Опрацьовані матеріали доступні за адресою: <http://www.artsrn.ualberta.ca/EnglishAudio/>. Я також мала можливість співпрацювати з українцями в Казахстані. Завдяки фонду імені Кулів і підтримці Алевтини Цветкової з Павлодарського державного університету імені Султанмахмута Торайгірова мені вдалося зібрати дані про обряди в селах на північному сході Казахстану. Цей район мене цікавив, оскільки українці оселилися там майже у той самий час, коли вони вперше почали переїжджати до Канади. Реформи Столипіна спричинили значну еміграцію з України до Павлодарського краю, про що свідчать назви таких сіл, як Красилівка, Соснівка і Хмельницьке. Звукові файли з цих досліджень також було проіндексовано відповідно до чинної системи, їх можна прослухати на сайті <http://www.artsrn.ualberta.ca/KazakhstanAudio/>.

Успіх і важливість проекту “Sanctuary” принесли йому ще більшу увагу, і нам вдалося отримати грантову підтримку від фундації Killam (Killam Foundation), Фонду спадщини Альберти (Alberta Heritage Foundation), Спадщини Саскачевану (Saskatchewan Heritage) та інших інституцій. Найбільший вплив на нашу роботу мав грант з KIAS (Kule Institute for Advanced Study, укр. – Інститут провідних досліджень імені Кулів). Цей грант підтримував польові дослідження і оплачував роботу помічників-аспірантів. Він також передбачав підтримку від бібліотеки Альберти. За згодою з бібліотечною системою Альберти наша робота стане частиною колекції Peel’s Prairie Provinces<sup>1</sup> і буде постійно зберігатися в бібліотеці. Оскільки ми співпрацюємо з бібліотекою, ми вирішили використовувати іншу систему кодування для наших даних. Фотографії, які зробили мої колеги Д.-П. Хімка і Ф. Свиріпа, стануть частиною однієї бази даних, попередня версія якої вже створена. Мої звукові файли стануть частиною іншої бази даних, і саме на цю роботу я хочу звернути тут увагу.

Співробітники бібліотеки підтримали мою мету зробити мої дані широкодоступними. Тому Шерон Фарнел (Sharon Farnel) й інші працівники бібліотеки запропонували перетворити нашу внутрішню систему індексування в анотовання, використовуючи метадані з Етнографічного тезауруса Американського фольклорного товариства (The American Folklore Society Ethnographic Thesaurus): <http://id.loc.gov/vocabulary/ethnographicTerms.html>. Цей Тезаурус є міжнародним стандартом і його використання зробило б мої записи легкодоступними у всьому світі; це також дало б змогу зацікавити мою роботою дослідників українського фольклору і науковців, які займаються вивченням фольклору інших регіонів світу. Перекодування моїх звукових файлів, навіть тих, які вже були проіндексовані, виявилось нелегким завданням. Категорії Тезауруса настільки відрізнялися від наших внутрішніх метаданих, що ми вирішили почати кодування ще не оброблених звукових файлів, плануючи пізніше повернутися до попередніх файлів і перекодувати їх. На цей момент, використовуючи нову систему – Етнографічний тезаурус, ми опрацювали записи, зроблені в 2013, 2014 і 2015 роках, тобто 79 годин аудіо.

<sup>1</sup> Peel’s Prairie Provinces – ресурс, призначений для надання допомоги науковцям, студентам і дослідникам різних напрямів, які вивчають історію Західної Канади і культуру канадських прерій. – *Примітка перекладача.*

Етнографічний тезаурус був створений Американським фольклорним товариством у Сполучених Штатах. Відтак він орієнтований на США не тільки з огляду на обрані групи метаданих, але і з погляду класифікації категорій. Для нашої роботи використання Тезауруса було проблематичним з двох причин. Однією з проблем є те, що категорії, які становлять інтерес для українців, часто вбудовані в систему кодування на кілька рівнів униз. Іншою – те, що терміни, які традиційно використовують українські вчені і самі ж українці, або взагалі не з'являються в Тезаурусі, або ж закодовані з використанням термінології, яка орієнтована на користувачів країн Західної Європи й Америки: тут часто вжито такі терміни для свят й інших обрядових реалій, які українці не використовують і яких навіть можуть не знати. Є також певні структурні проблеми, пов'язані з Тезаурусом, які будуть коротко згадані після дискусії, що стосується безпосередньо українців.

Проблему “глибокого” вбудування актуальної для української науки термінології в систему кодування Тезауруса можна побачити в наступних двох прикладах. Церкви надзвичайно важливі для українців Канади і були в центрі уваги нашої роботи. Однак у Тезаурусі термін “*церква*” виступає останнім у довгому ряді категорій, а саме: матеріальна культура > об'єкти > архітектура > будівлі > релігійні споруди > церкви. “Церква” є останньою в ланцюжку шести термінів. Така сама ситуація з Пасхою, яка є основним святом для українців Канади й українців загалом, але є восьмою в ряді термінів, серед яких – ритуали > календарні обряди > свята > релігійні свята > Великий піст > Страсний тиждень > Великдень. Хоча пошукова система приведе користувача безпосередньо до потрібної категорії, структура категорій не виражає взаємопов'язаних рівнів і може вплинути на те, що інформація про Україну здаватиметься менш важливою.

Іншою проблемою є відсутність підкатегорій, важливих для українського матеріалу. Їжа є центральною частиною всіх обрядів; деяка їжа, наприклад, *коровай* (який є елементом весілля), *паска* і *бабка*, (елементи Пасхи), і *калач* (що побуває як на Різдва, так і під час похоронних обрядів і поминань померлих), є характерно українськими та ключовими для українських традицій і вірувань. Жодного з цих термінів немає в Тезаурусі. Що важливіше, обрядові страви є найнижчою категорією в групі метаданих “їжа”. Як уже зазначалося, ритуальні страви є частиною усіх українських обрядів, і користувач, який виконує пошук у цій категорії, отримає аж 193 пошукові одиниці лише серед звукових файлів, які ми опрацювали на цей час. Коли пошук видає таку велику кількість файлів, то користувач потребує повторного пошуку, щоби знайти потрібний матеріал. Натомість у Тезаурусі користувач не лише не зможе дізнатися, що *коровай* використовується як весільний атрибут у певному регіоні Канади, він також не зможе розрізнити застілля, яке відбувається під час весілля, від частувань, що є частиною обряду Маланки. Користувач матиме два варіанти. Перший – натиснути на кожну зі 193 позицій, які висвітляться у результаті пошуку обрядових страв, і кожну з них перевірити, щоб з'ясувати, які види страв обговорюються і до якого обряду страва належить. Інший варіант полягає в пошуку за ритуалом, а не за стравами. Оскільки Етнографічний тезаурус орієнтований на США, упорядкування категорій залежать від сучасних тенденцій в американській науці. Традиції, пов'язані з їжею, є популярним підрозділом фольклору, і для дослідників цих традицій те, чи страва

є обрядовою, чи належить до іншої категорії їжі, буде важливішим, ніж обряд, у якому вона використовується. Для українців це може бути навпаки. Наукові інтереси Заходу також проявилися у розрізненні ікон та іконопису в Тезаурусі. Ікони належать до ритуальних і обрядових предметів, а іконопис – до мистецтва. Такий розподіл не є корисним для наших даних і спричинений тим, що американські фольклористи вірогідніше будуть зацікавлені або в мистецтві, або в обряді, але не в поєднанні цих двох сфер.

Проблема відсутності української термінології є серйозною. Дні святих мають свою категорію в Етнографічному тезаурусі, але багатьох важливих для українців святих немає в списку. До них належать святі, на честь яких були названі церкви, – святі Петро і Павло та святий Миколай. У преріях однією з найважливіших подій є власне богослужіння, яке проводиться в день святого, на честь якого названа церква. У православних це називається *Храм*, а у католиків східного обряду – *Празник*. Терміни “Храм” і “Празник” не є частиною Тезауруса і не можуть використовуватися як альтернатива до дня святих. *Вивід матері* (по-народному: виводина) – обряд, який іноді поєднується з хрещенням, а іноді виконується окремо, не з’являється в Тезаурусі. Тезаурус подає категорію “хрещення”, але вона не має жодних підкатегорій. Є й безліч інших прикладів. Майже всі українці Канади пам’ятають, як дітьми ходили до будинків сусідів і родичів, щоб *посіяти*, – обряд, який охоплює кидання зерна і декламацію невеликої строфи з побажаннями благополуччя й удачі дому, який відвідують. Цієї події немає серед категорій Тезауруса, а найближчою категорією, яку ми можемо відшукати, є Новий рік.

Деякі важливі події кодуються термінами, яких українці не використовують і які, ймовірно, не будуть шукати під час використання нашої бази даних. Іншими словами, є певні календарні свята, які в уродженців Заходу мають одну назву, а в українців Канади – іншу. Наприклад, українці використовують термін “*Вербна неділя*” на позначення неділі перед Великоднем і мають у традиції веснянки та приказки, пов’язані з вербою, яка характеризує цей день. Терміна “Вербна неділя” (англ. – Pussy Willow Sunday) немає в Етнографічному тезаурусі, а неділя перед Великоднем усього лише позначена як “Пальмова неділя” (англ. – Palm Sunday). Деякі церкви католиків східного обряду використовують термін “Пальмова неділя”, тож термін, котрий обирає респондент, може свідчити про релігійну приналежність. Тому для нашого дослідження є важливим розрізнення уживання термінів Вербна неділя і Пальмова неділя. *День Святої Трійці* (англ. – Trinity Sunday), важливе для українців літнє свято, на честь якого було теж названо багато церков, не зазначений окремим терміном в Етнографічному тезаурусі. Натомість у Тезаурусі подано – “*П’ятидесятниця*” (англ. – Pentecost) і не вказано жодного синоніма назви цього свята. Українці часто називають його “*Зелені свята*”, а термін “П’ятидесятниця” розпізнають не одразу. Ритуал освячення води, який відбувається через дванадцять днів після Різдва і який українці називають *Йорданом*, закодований в Тезаурусі як “Епіфанія” (англ. – Epiphany). Якщо українці й використовують для цього свята інший термін, ніж “Йордан”, вони називатимуть його радше “*Богоявленням*” (англ. – Theophany). Зазначений у Тезаурусі термін “Епіфанія” українці використовують рідко і його пошук приведе неукраїнців до українського матеріалу, а українцям, які шукатимуть це свято під назвою “Йор-

дан” – за терміном, який їм відомий, – може видатися, що інформація про нього недоступна в нашій базі даних.

Оскільки ми прагнемо використовувати Етнографічний тезаурус і в такий спосіб привернути увагу світової спільноти до українського матеріалу, ми застосували двостороннє вирішення зазначених вище проблем. Перше рішення – додати специфічно українські категорії: на сьогодні ми визначили і додали 24 категорії такого типу. Тобто до нашої групи метаданих ми додали всі специфічно українські категорії, перелічені вище, і декілька додаткових. Тепер у нас є метадані термінів “Йордан”, “Вербна неділя”, “День Святої Трійці”, “сіяти”, і всіх святих, на честь яких названі церкви і яких вшановують на відповідному “Храмі” або “Празнику”. Для таких термінів, як “Вербна неділя”, ми використали більш західне позначення “Пальмова неділя”, якщо цей термін використав сам респондент запису. Якщо ж мовець вибрав для позначення цього свята термін “Вербна неділя”, ми його теж зазначили.

Серед даних, які ми досі опрацювали, налічується 2 223 індексовані секції і 193 категорії. З них 24 категорії є винятково українськими і охоплюють 349 індексованих секцій. Інші дані (ті, які збігаються з Етнографічним тезаурусом Американського фольклорного товариства) містять 1 874 індексовані секції в 169 категоріях. Ми переконані, що визначені нами 24 нові категорії є актуальними і дійсно відображають обрядову практику канадських українців. Наша впевненість ґрунтується на тому, що наш доробок показує високий ступінь повторюваності. З цих нових категорій 13 було знайдено в 10 або більше секціях, а 9 категорій – у 20 або більше секціях. Йордан, наприклад, був згаданий в інтерв’ю 46 разів.

Визначення цих 24 термінів, що вживаються канадськими українцями, означає, що у нас є можливість розширити Етнографічний тезаурус на 24 групи метаданих. Ми будемо проводити зустрічі за круглим столом на засіданні ASEES (англ. – Association for Slavic, East European and Eurasian Studies) – Асоціації слов’янських, східноєвропейських та євразійських досліджень, заплановану на листопад 2016 року у Вашингтоні. На зустрічі ми обговоримо характерні слов’янські етнографічні категорії і підготуємо проект звернення до Американського фольклорного товариства з пропозицією розширити Етнографічний тезаурус і внести до його складу визначену нами українську й іншу слов’янську термінологію. Етнографічний тезаурус додав категорії з інших культур, які практикують свої традиції в Північній Америці. Одним із прикладів є індістське календарне свято Дівалі. Ми докладемо зусилля, щоб характерні українські і більш загальні слов’янські категорії були долучені до цієї міжнародної системи метаданих. Ми співпрацюємо з Етнографічним тезаурусом трохи більше року і помітили, що він постійно змінюється: змінюються як самі категорії, так і зв’язки між групами метаданих. Схоже, це особливо вдалий час, щоб запропонувати додати нові українські категорії.

Нашим іншим підходом буде додати дві пошукові опції на головній сторінці нашої бази даних. Першою – буде випадне меню, що використовуватиме Етнографічний тезаурус і приваблюватиме користувачів, які не знайомі з українською культурою, але хочуть більше дізнатися про українські обряди і фольклор або бажають знайти український матеріал, щоб порівняти його з культурою, яку вони

досліджують. Цей вид пошуку допомагатиме тим, хто не є фаховими українцями, а хоче порівняти українські явища з географічними та культурними регіонами свого дослідження. Дерево (розгалужену систему), на якому базуватиметься цей пошуковий портал, можна побачити на нашому веб-сайті <http://www.artsmn.ualberta.ca/orodrigu/tree/>. Інше випадне меню на головній сторінці матиме значно менше метаданих термінів і буде орієнтоване на українських користувачів. Такі терміни, як “Йордан”, “День Святої Трійці” і “Вивід матері”, з’являться одразу і не будуть вбудовані у велику кількість категорій, як це характерно для Етнографічного тезауруса. Я також розглядаю можливість написати вступну статтю про культуру канадських українців. Отож, ми повинні створити два випадні меню і випробувати їх, щоб побачити, які додаткові вказівки будуть необхідні (наприклад, вступна стаття).

Кодування українсько-канадських даних за системою метаданих, непристосованою до цього типу матеріалу, є складним завданням. Але це те важке завдання, яке потрібно виконати. З початку своєї етнографічної діяльності я прагнула зробити українську культуру настільки широковідомою, наскільки це тільки можливо. Мої студенти у Вірджинському університеті не були українського походження. Їхня робота з моїм українським веб-сайтом з метою пізнати принципи фольклористичної діяльності сприяла підвищенню їхніх знань про Україну та українську культуру. База даних з аудіозаписами, зробленими в Україні, відкрила публіци доступ до матеріалів українською мовою. Їх використовували переважно дослідники, але сайт був також корисним для північноамериканців, які знають розмовну українську, в тому числі і тих, хто не вміє писати українською. Деякі з моїх українсько-канадських звукових файлів англійською мовою вже доступні з використанням нашого старого сайту зі старою системою пошуку. Нова база даних, створена на основі Етнографічного тезауруса, стане доступною найближчим часом і охопить якомога ширшу аудиторію. Ми творимо цей сайт із наміром, щоб він також надавав легкий доступ користувачам з української спільноти. Вони будуть обслуговуватися в рядку пошуку, орієнтованому спеціально на українські інтереси. Виходячи за межі власної бази даних і працюючи, щоб здобути вплив на міжнародному рівні, ми прагнутимемо розширити Етнографічний тезаурус шляхом залучення у нього українських категорій і таким способом підвищити обізнаність з українською культурою всіх, хто використовує Тезаурус з будь-якою метою, в тому числі з метою кодування власного матеріалу. Навіть ті люди, які ніколи не користувалися нашою базою даних, але є фольклористами і використовують Тезаурус, щоб класифікувати власний матеріал, ознайомляться з українським фольклором завдяки його долученню до міжнародно визнаної системи.

#### Список використаної літератури

1. *Kononenko Natalie*. How God Paired Men and Women: Stories and Religious Revival in Post-Soviet Ukraine // *Canadian-American Slavic Studies*. – 2010. – Vol. 44. – № 1–2. – P. 105–133.
2. *Kononenko Natalie*. *Ukrainian Minstrels: And the Blind Shall Sing*. – Armonk, New York, London: M.E. Sharpe, 1998. – XVI, 360 p., ill.



3. *Бріцина Олександра, Головаха Інна*. Прозовий фольклор села Плоске на Чернігівщині: Тексти та розвідки. – Київ: Інститут мистецтвознавства, фольклористики та етнології імені М. Т. Рильського НАН України, 2004. – 584 с.

**BUILDING UKRAINIAN FOLKLORE DATABASES:  
COMMUNICATING INFORMATION  
TO AN INTERNATIONAL AUDIENCE**

**Natalie KONONENKO**

*Peter & Doris Kule Centre for Ukrainian & Canadian Folklore,  
Modern Languages and Cultural Studies, Faculty of Arts, University of Alberta  
441-C Old Arts & Convocation Hall,  
tel.: 780-492-6810, e-mail: natalie.kononenko@ualberta.ca*

Kononenko and her team of programmers and researchers began digital presentation of Ukrainian folklore field recordings in 2000. This article describes the author's discovery of the relationship between the possibilities offered by online data management and the needs of users. It details her efforts to find a way to best present Ukrainian folklore to both researchers and the general public. Digital technologies offer unprecedented opportunities for the publication of sound, capturing nuances that cannot be represented in written form (transcription). Because of the advantages that the online world offers in terms of audio, the article then focuses on the history of Kononenko's work with sound file databases. It describes Ukrainian Folklore Audio, a website where approximately 200 hours of folklore recordings made in Ukraine are available for public and researcher use. The special virtue of this site is that it was among the first databases which allowed the user could go directly to the point in a recording where a particular topic was discussed. The article describes the process of building this site and the challenges encountered by researchers and programmers who were working in uncharted digital territory.

Kononenko is now collecting recordings describing Ukrainian ritual practices in Canada. These are encoded according to the International Ethnographic Thesaurus, a set of standard metadata that has, to date, not included much material on Ukraine. The work of Kononenko and her team promises to expand the Ethnographic Thesaurus to include 24 Ukrainian-specific items. The Sanctuary Project Audio Files will be housed in the Peel's Prairie Provinces Collection of the University of Alberta Library. The website is currently being testing and should soon open for public use.

*Key words:* folklore fieldwork, digital technologies, Ukrainian folklore in Canada, The Ukrainian Folklore Sound Recordings website, International Ethnographic Thesaurus.