Proposal to revise existing ALA/LC Macedonian Romanization Table

This is a proposal to revise the ALA-LC Romanization table for the Macedonian language. The revision is needed to distinguish the romanization of:

- the Macedonian letter "**Ѕ/ѕ**" from the two-letter sequence "**Дз/дз**"—both currently romanized with the Latin-alphabet digraph "**DZ/dz**"
- the Macedonian letter "**Џ/џ**" from the two-letter sequence "**Дж/дж** "—both currently romanized with the Latin-alphabet digraph "**DŽ/dž**"

Explanation:

The current Macedonian Romanization table does not distinguish between the digraphs and the two-letter sequences, and catalogers performing data entry use the sequence of Latin letters "dz" to transcribe both "дз" and "ѕ," and "dž" to transcribe both "дж" and "џ." According to the table, whenever the sequences "dz" and "dž" occur in the romanized form of a Macedonian word, they are to be interpreted as "Ѕ/ѕ" or "Џ/џ", respectively. This is problematic because the foregoing interpretation of the sequences "dz" and "dž" is not the only possible interpretation of those sequences; indeed, the sequences "дз" and "дж" in Macedonian are common and appear in several contexts:

1. When a prefix ending in "-д," such as над- (*nad-*), од- (*od-*), под- (*pod-*), пред- (*pred-*) is attached to a root starting with a "з" or "ж". Examples include several dozen words and their derivatives, including common vocabulary such as надзор (*nadzor*) 'supervision', одзади (*odzadi*) 'behind', подземи (*podzemi*) 'underground', предзнак (*predznak*) 'omen', одживен (*odživen*) 'obsolete', надживее (*nadživee*) 'to outlive', поджегне (*podžegne*) 'to sting, offend'.
2. In foreign loans and foreign proper names, for example рендзина (*rendzina*) 'rendzina' (a type of soil), Виндзор (*Vindzor*) 'Windsor', Джузепе Гарибалди (*Džuzepe Garibaldi*) 'Giuseppe Garibaldi'.
3. In acronyms, especially since државен завод (*državen zavod*) 'state institution' is a common phrase in names of governmental organizations, for example ДЗС (Државен завод за статистика) (*Državen zavod za statistika*, State Statistics Office), ДЗИС (Државен завод за индустриска сопственост) (*Državen zavod za industriska sopstvenost*, State Office of Industrial Property).

According to the current Macedonian Romanization table, these two-letter sequences should be romanized as "dz" and "dž". However, because "dz" and "dž" are already reserved as romanizations of the characters "Ѕ/ѕ" and "Џ/џ", they cannot be used as romanizations of the letter sequences "дз" and "дж". There is thus no way in which to

represent "дз" and "дж" unambiguously in romanized form, and the current system does not adhere to the principle described in section III.2.b of the LC's Procedural Guidelines for Proposed New or Revised Romanization Tables: "...Latin script equivalents should allow for the reversal of Romanization as systematically as possible, without the application of special algorithms or contextual tests." One practical consequence of this is that any macro for Roman-to-vernacular script conversion will automatically render "DZ/dz" and "DŽ/dž" as "S/s"and "Џ/џ", respectively, even in contexts where the romanized form "dz" or "dž" is intended to represent "дз" or "дж".

Example:

Original vernacular script field: Самоуправување и облици на на**дз**ор во работните организации

Romanized title: Samoupravuvanje i oblici na na**dz**or vo rabotnite organizacii.

Incorrect vernacular script field resulting from applying macro: Самоуправување и облици на на**s**ор во работните организации.

Research:

In the past, the Committee has made proposals for similar linguistic cases, for example, for the Belarusian Romanization Table, that suggested only a minor change—the addition of notes in the Romanization table—and required manual review on the part of catalogers. Years of observation of how this change has worked in the real cataloging environment, however, have shown that this approach only works partially, when manual review by an experienced cataloger is available. In many other instances, when transliteration is run through the macro but a cataloger isn't knowledgeable enough to perform a manual review, the problem of incorrect transliteration for Belarusian materials persists.

In order to better understand how a change in Macedonian Romanization might affect legacy bibliographic data, the Committee approached the OCLC Metadata Quality Control Section with a request to run a query for all instances of words in Macedonian-language records with occurrences of the two letter sequence "dz", including such occurrences in the middle of a word. OCLC staff used a tool internal to OCLC and identified 1,657 bibliographic records for Macedonian materials in the OCLC database that included occurrences of "dz" (with or without diacritics). All 1,657 records would need some manual review; approximately 30% from this list would require changes in Romanization.

The Committee was encouraged by the relatively small number of records that would need manual review and decided to proceed with the proposal of a substantive change to the current Romanization Table. Numerous options for novel ways to represent both "S/s" and "Џ/џ" were discussed, with feedback solicited from outside the Committee as well, including from catalogers in the Library of Congress' Southeast Europe Section. The outcome of these discussions is presented below. Members of the Committee are willing to undertake the project of manually reviewing all records from the list that was sent by the OCLC Quality Control team if the proposal is approved. Since the LC catalog is separate from OCLC, we hope that some arrangements could be made as to how to deal with affected records in the LC database as well.

Proposed solutions:

1.   Change Romanization for the Macedonian letter "S/s" from "DZ/dz" to "D͡Z/ d͡z" (adding combining ligature marks).

The Committee felt that it was imperative to preserve the underlying sequence of "dz" for the representation of the Macedonian letter "s." Because most public search systems ignore diacritics, retaining "dz" and simply adding diacritics would allow patrons to continue using library systems as they are accustomed; were a single-letter solution such as "ż" to be introduced, the Slavic librarian community would have to undertake a massive international outreach campaign to alert researchers to modify their search practices. The use of the combining ligature marks is consistent with standards for other Slavic languages, such as "t͡s" for "ц" in Bulgarian and other languages, and it conveys clearly the single affricate phoneme represented by the Macedonian letter "s." Moreover, the Macedonian Academy of Arts and Sciences prescribes "dz" as its standard for Romanization of "s," so the addition of combining ligature marks can be seen as the least "disruptive" modification that would still solve the problem of ambiguity between Cyrillic "s" and "дз."

2. Change Romanization for the Macedonian letter "Џ/џ" from "Dž/dž" to "D͡Ž/ d͡ž" (adding combining ligature marks).

As with the other proposed change, this modification would preserve the underlying sequence of "dž," allowing experienced patrons to continue using front-end search systems without disruption, and, again, deviating only superficially from the transliteration standard of the Macedonian Academy of Arts and Sciences, which prescribes "dž." It is true that the combination of ligature marks and the hachek diacritic is not ideal visually or technically; however, the use of multiple diacritics is not without precedent in the ALA-LC Romanization Tables, and online displays such as the LC's

OPAC and its Voyager cataloging interface render this combination of characters successfully.

Both of these modifications would significantly improve automated machine transliteration and the unambiguous interpretation of Latin-only bibliographic records.

Sources:

De Bray, Reginald, et al. *Routledge Macedonian-English Dictionary.* London/New York: Routledge, 1998.

Ristovski, Blaže. *Makedonska Enciklopedija.* Skopje: MANU, 2009.

The proposed revised Macedonian Romanization Table is provided in an attached PDF.

Proposal compiled by working group members:
Larisa Walsh (University of Chicago, chair)
Kirill Tolpygo (University of North Carolina, Chapel Hill)
Thomas Dousa (University of Chicago)
Cammeron Girvin (Library of Congress)
Geoff Husic (University of Kansas) (consultant)